PART II

THE ASSESSMENT CONFERENCE AND CONCLUSIONS AS TO THE STATE OF THE ARTS OF THE CVM

VII. THE ASSESSMENT CONFERENCE: OVERVIEW

On July 2, 1984, a conference "Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Valuation Method" was held at the Hyatt Palo Alto Hotel in Palo Alto, California; some eighty professional researchers with interests in the public goods valuation issue attended the conference. Most conference participants received Part I of this book (Chapters I - VI), or an Executive Summary of Part I, several weeks prior to the conference.

The conference format was as follows. During the morning session, papers were presented by Professors Richard Bishop (University of Wisconsin), A. Myrick Freeman (Bowdoin College), Alan Randall (University of Kentucky) and V. Kerry Smith (Vanderbilt University). Papers presented by these four scholars are given below in Chapters VIII - XI. Generally, these authors address two major issues in their papers: their critical review of this books' Part I (Chapters I - VI), and their individual assessment of the state of the arts of the CVM. The afternoon session was devoted to comments offered by a Review Panel. Members of the Review Panel were: Kenneth Arrow (Stanford University), Daniel Kahneman (University of British Columbia), Sherwin Rosen (University of Chicago) and Vernon Smith (University of Arizona). Based on their pre-conference reading of this books' Part I and the four papers presented in the morning session, comments by the Review Panel were focused on each Panel member's assessment of the strengths and weaknesses of the CVM as a means of estimating social benefits attributable to environmental (and public) goods. Comments by the Review Panel are given below in Chapter XII.

Thus the following five chapters review the results from the assessment conference and provide the reader with diverse views concerning first, the authors' analysis of the CVM given in Part I and, second, the strengths, weaknesses and promise of the CVM. Conference results presented in these five chapters serve to set the stage for the ultimate task of this book: the offering of conclusions concerning the state of the arts of the CVM. The development of such conclusions is the topic of Chapter XIII given below.

VIII. THE POSSIBILITY OF SATISFACTORY BENEFIT ESTIMATION WITH CONTINGENT MARKETS

By
Alan Randall*
Department of Economics
University of Kentucky

Skepticism about the contingent valuation method (CVM) has always focused on value data quality. It has long been clear that, if the value data can be trusted, these data (unlike the data used in weak complementarity and hedonic price theory approaches) can be directly interpreted as estimates of welfare change consistent with accepted economic theory (Bradford 1970, Randall *et al*. 1974, and Brookshire *et al*. 1980). However, CVM data are self-reported by participants in interaction with a researcher or his/her representatives. This gives rise to obvious concern that various self-reporting biases, and other biases inadvertently introduced by the research design and/or the interaction between researcher and participant, may be endemic to CVM.

On the other hand, a quite considerable body of empirical evidence can be broadly interpreted as supportive of CVM. True, unexpected and perplexing results occur from time to time. Nevertheless, the broad thrust of the empirical evidence is to corroborate CVM findings. This was my perception prior to reading the Cummings *et al*. "State of the Art" document, and that document tends to reinforce my prior perception.

The Cummings *et al*. document also reinforces another of my prior perceptions: that the research approach toward investigating data quality in CVM has been skewed toward the empirical. In some cases, empirical experiments have been designed to address data quality issues. In others, data quality issues have been addressed *ad hoc*, as apparently anomalous results have seemed to require *ex post* interpretation. The net result has been the accumulation of a detailed taxonomy of "biases in CVM." One problem with this apprach has been a tendency to promulgate empirical laws on the basis of a few small-sample data sets. Another has been a rather widespread failure to critically scrutinize the notion of "bias" itself, to specify what conditions are sufficient for an unexpected result to be correctly interpreted as attributable to bias inherent in the data collection method.

Resolution of controversies about data quality in CVM seems sure to require a formal theory of the behavior of participants in CV exercises. John Hoehn and I (manuscript) have recently developed the rudiments of one such theory. In this paper, I will outline the intuition behind this theory and suggest its usefulness in (1) predicting the direction of any deviations of CVM-reported benefit values from optimally-formulated values, (2) explaining certain empirical results previously though anomalous, and (3) identifying procedures to improve the accuracy of CVM and render the direction of the remaining inaccuracy more predictable. I hope the following discourse will achieve two objectives: to illuminate the data quality issues in ways that empirical evidence alone cannot do; and to demonstrate, by methodological example and through its results, that CVM is a progressive research program in the sense of Lakatos (1970).

A. A THEORY OF CVM-PARTICIPANT BEHAVIOR

Many of the purported biases in contingent valuation seem to be rather simple concerns that can be avoided or minimized through careful attention research design, sampling, and administration of the experiment or survey. Two concerns that are genuinely interesting are: (1) individuals may behave strategically, misreporting their "true" valuations in order to benefit themselves by influencing the outcome of policy research; and (2) individuals may treat the whole exercise as inconsequential, and thus devote little effort to the introspection that is necessary to discover what one's "true" valuation really is. Hoehn and I (manuscript) address these two concerns, assuming a rational, self-seeking respondent and using simple economic-theoretic models to predict her behavior in a CVM setting.

1. Value Formation

First, assume the individual -- an experimental subject or survey respondent-- believes the results of the valuation exercise will influence policy. It is not essential to believe that it will be decisive; influential is enough. 1/ Assume also that the individual perceives that she is a member of a sample of citizens participating in the exercise. Does she "take it seriously?" It is reasonable to assume she will take it at least as seriously as voting in elections or participating in a political poll (where, again, her influence is magnified because she is a member of a sample chosen to represent a larger population). Since policy choices are more focused and more precisely specified in CVM than in elections and political polls, it is possible that participants may feel that CVM offers them an unusually favorable opportunity to influence policy choice.

Now, assume that formulating ("figuring out") her WTP/WTA for specified changes in Q (or, even more difficult, specifying her total value curve) is not so simple a task that it can be accomplished instantaneously and costlessly. The choices offered in the contingent market will seldom be familiar and routine, even with the best research design. There will be a positive relationship between the effort she invests in value formulation and the precision of the value at which she arrives. If the value formulation task is very difficult and/or the individual limits the effort she invests therein, she may solve the value formulation problem imcompletely or imprecisely.

This places in perspective the difference between contingent markets and "real" markets. First, the goods offered in contingent markets are not always familiar, and individuals may not associate these particular goods with trading possibilities. Nevertheless, unfamiliar goods are often introduced in "real" markets and, especially, in market experiments. So, this distinction between "real" and contingent markets is, if anything, a matter of degree. Second, the penalty for a wrong decision may be substantial in "real" markets: your money is gone and you are left with some purchase that has disappointed you. There is, however, a penalty for a wrong decision in a contingent market: one's opportunity to influence policy is wasted or misused and one's chances of facing a less-preferred policy environment are accordingly increased. Again, the distinction between "real" and contingent markets is, if anything, a matter of degree. Sub-optimal individual decision making can be expected in both kinds of market, but may be more prevalent in contingent markets.

If value formulation is imperfect in contingent markets, the formulated

values would include some error. Can we identify the direction of that error? It turns out that if valuation is performed in the Hicksian compensating framework (i.e., WTP for increments in Q and WTA for decrements), imperfect value formulation would lead to understatement of WTP and overstatement of WTA.

The intuitive explanation of this result is as follows. In order to formulate her WTP, the participant must first solve the problem: minimize expenditure subject to utility constrained at the initial level. Imperfect solution of that problem can have only one kind of outcome, the identification of some expenditure larger than the minimum. This overestimation of minimum expenditure must lead the participant to underestimate her compensating surplus, WTP. Thus, any error in formulating WTP in a compensating framework would lead to its understatement. 2/ This line of reasoning further suggests that WTP is nondecreasing in the time (and by extension, other resources) allocated to solving the value formulation problem.

To summarize, incomplete value formulation in a Hicksian compensating context tends to understate WTP (and overstate WTA); and the magnitude of the error is nonincreasing as more effort is invested in the value formulation process.

2. Value Reporting

Now, assume the individual is not above strategic behavior, which we define as reporting something other than one's formulated value in order to influence policy in one's favor. Some participants would reject this kind of behavior on moral grounds, while others would recognize that strategic behavior is itself resource-consuming and decide not to use resources that way. Nevertheless, it is surely prudent to consider what kind of effect those who choose to attempt a strategic response might have on reported contingent valuation results.

To identify optimal strategies for participant, we must first specify the incentives that they face. For simplicity, assume that $U = U(Q,Y)$, where Q is a nonmarketed amenity and Y is a numeraire consisting of "all other goods." Assume the individual gains positive utility from both Q and Y. In other words, she likes Q and does not like taxes or payments that would reduce her disposable income for purchasing other goods. The key issue, then, is how her participation in the exercise is likely to influence (1) the chances that a policy to increase Q will be implemented and (2) her disposable income, if the policy is enacted. One can model a variety of alternative contingent markets to examine how their structure affects these things. Here we outline some of these models for WTP; the arguments are analogous for WTA, where the effects are usually similar but of opposite sign.

We can dispose quickly of two rather obvious cases.

a). The agency will provide the increment in Q without regard to the outcome of the benefit cost analysis. The researcher will collect stated WTP from each participant at the end of the exercise. However, Q is nonexclusive and participants will enjoy the increment in Q regardless of their reported (and paid) WTP. Strategizing respondents would report zero or very low values for WTP.

b). The agency will provide the increment in Q if and only if the estimated benefits for the affected population exceed the costs. The researcher never collects the stated WTP, and nor does anyone else. The participant is forever immune from bearing any of the costs. Strategizing

respondents would state high values for WTP in order to increase the probability of implementing the policy.

These cases can immediately be dismissed since they are quite false representations of the policy environment. Case (a) is of some interest in experimental economics, as the case most likely to elicit free-rider behavior. However, it is not common policy practice to implement proposals independently of benefits and costs, and to finance them through contributions determined by self-reported WTP. Case (b) has some appeal on the surface, since in BCA practice the researcher seldom collects WTP. However, a deeper analysis suggests that participants realize that if the exercise is to affect policy they will eventually pay -- usually through some combination of user fees, higher taxes, and higher prices -- for increments in Q. The assumption that the participant is forever immune from contributing toward the costs of policy is untenable.

Cases (a) and (b) share and interesting characteristic: they deviate from the policy choice model in that the respondent is not attempting simultaneously to influence Q and Y. In case (a), Q is given and the respondent has only to maximize Y. In case (b), Y is not at issue and a Q-loving respondent has only to maximize the probability that is provided.

More relevant models of the incentives influencing behavior in contingent markets include the following cases:

c). The proposal is implemented if the estimated benefits exceed the costs, and citizens pay in proportion to stated WTP. In this case the respondent influences her payment in the event of policy implementation and the probability of implementation. She faces uncertainty about project costs and about the aggregate reported benefits.

d). The proposal is implemented if the estimated benefits exceed the costs, and citizens pay their share of the costs, as determined by some pre-specified rule. In this case the respondent influences the probability that policy is implemented and payment exacted. She faces uncertainty about project costs (and thus the size of individual cost shares) and about reported aggregate benefits.

e). The proposal is implemented if a plurality of citizens approves it, given information on the individual payment to be exacted. Since the expression of approval is condition on a stated level of payment, the level of payment can be varied and the question of approval reiterated. The respondent is uncertain about how others will 'vote', which provides incentive for participation. Uncertainty about the true level of policy costs is neither essential nor damaging to the incentive properties of this decision rule.

In each case the participant who likes Q but dislikes bearing additional expenses must devise a strategy designed to increase the expectation that the policy is implemented but, ceteris paribus, reduce the expected cost she will bear.

Optimal reporting strategies for cases (c) through (e) are:

c.) Report WTP equal to or less than one's formulated WTP. Optimal reporting strategy is related to sample size. Generally it is best to report WTP approaching one's formulated WTP, if one believes the sample is small. With very large samples the tendency toward free-riding is stronger if the CV exercise is treated as a one-shot game; if it is treated as one play in a repeated game with an indefinite end-period, the cooperative strategy of truthful reporting may emerge.

d). If one suspects one's formulated WTP is quite different from that of other citizens, exaggerate the difference so as to shift the sample mean reported WTP nearer to one's own formulated WTP. If one expects one's WTP is a little higher than the mean, report a value still higher; likewise, if one's WTP is likely to be lower than the mean, report a value still lower. Again, if the CV exercise is treated as one play in a repeated game, truth-telling may be prevalent.

e). No strategy is individually preferred to truth-telling. If the stated individual cost is lower than one's formulated WTP, it is optimal to report approval; if one's WTP is lower than the stated cost, it is individually optimal to report disapproval.

What effect would these individually optimal strategies have on estimated benefits of increasing the level of Q? In case (c) there may be a tendency to underestimate benefits. In case (d) the variance of individual WTP may be increased, widening the confidence interval around estimated benefits. If reported WTP is limited to a minimum of zero but has no upper limit, mean reported WTP might be biased upward. However, there are statistical methods for dealing with this problem. If these methods are used, total estimated benefits would be unaffected by reporting strategies.

In case (e) there is no reporting bias. Note that in this case the results are expressed in terms of "number of participants expressing approval/disapproval of the proposal given a per capita cost of $____." These results are not immediately interpreted as WTP. All we know is that those who approve have formulated a WTP greater than the stated cost, while these who disapprove have formulated a WTP less than the cost. Nevertheless, all is not lost for the benefit cost analyst. If (1) the sample is subdivided and different subsamples respond to different stated costs and (2) the data are analyzed with appropriate statistica1 tools (e.g., logit analysis), valid benefit estimates can be obtained. An alternative approach is to repeat the "approve/disapprove" question with the same participant, stating different levels of individual cost. In that way te researcher could iteratively approach the participant's indifference point, while retaining the desired anti-strategic properties of the "majority vote" format.

3. Implications

This conceptual analysis of the participant's likely behavior in a contingent valuation exercise, in formulating and reporting her responses, has several implications; and these implications appear to have been corroborated in empirical applications.

First, while the incentives for careful decision making and truthful reporting of valuations are perhaps not as strong as in private goods markets, they are by no means absent in contingent valuation exercises. This suggests that carefully designed contingent valuation studies will collect a substantial body of serviceable value data. Economists have long recognized that private goods markets do not require, for their efficient functioning, that all participants make near-optimal decisions. Price-making at the margin is disproportionately influenced by arbitrageurs, and the mistake-prone are eliminated from the market. Public goods markets ("real" or contingent) do not have these characteristics. Thus a minority Of "dubious" value observations tends to persist in these markets. The earlier intuition of Randall et al. (1981) that empirical analysts focus on identifying the "solid core" of reliable observations, seems sound in

light of these considerations, CVM results, whether in the form of aggregate benefit estimates or tests for "bias," should not be overly influenced by a relatively few eccentric observations.

Second, for a fairly wide range of contingent market designs, we can be confident that any biases introduced in formulating and/or reporting WTP will have the effect of understating it. This applies to contingent markets based on Hicksian compensating measures of value, and assumes use of appropriate statistical analyses. Following Hoehn and Randall (manuspcript), we can define a satisfactory benefit cost estimator as one that correctly identifies all proposals that would not generate a potential Pareto-improvement (PPI) while correctly identifying at least a subset of those that would bring about PPIs. It follows that any BC estimator that reliably reports WTP (i.e., benefit) estimates no greater than their "true" values and WTA (i.e., costs) no less than their "true" values is satisfactory. 3/ Thus, we can identify a considerable class of CVM formats that are satisfactory BC estimators.

Third, contingent valuation formats come in considerable variety, and their performance characteristics will differ in ways that are, to some extent, predictable. Thus, the quality of contingent value data can be improved with careful attention to contingent market design. Use of Hicksian compensating value measures and referendum formats, as in case (e), are obvious ways to minimize bias in estimated benefits while ensuring that any remaining bias is toward understatement. Since strategic misstatement can be minimized or eliminated in this way, the commonly expressed fear -- that routine use of CVM to guide actual policy decisions would lead to rampant "strategic bias" -- seems misdirected. On the contrary, it seems desirable to emphasize the connection between CVM and policy decisions to enhance the incentives for careful value formulations.

Fourth -- since we have concluded that (i) a class of formats can be identified in which any inaccuracy would tend to understate WTP and overstate WTA and (ii) the divergence between WTP and WTA is nonincreaseing with value formulation inputs; and Hovis et al. (manuscript) have provided empirical evidence entirely consistent with our theoretical conclusions -- I see not great merit in the Cummings et al. recommendation that the profession abandon attempts to measure WTA with CVM. 4/

Finally, the identification of a class of satisfactory benefit estimators that use CVM data is not an invitation to complacency. Our definition of satisfactory BC indicators permits adverse evaluations of some proposals that would generate PPIs. Obviously, it would be desirable to continue refining our understanding of CVM to identify approaches to reduce the frequency of this kind of misevaluation.

B. "INFORMATION BIAS" AND POLICY EVALUATION

In 1983 I wrote (with John Hoehn and David Brookshire) some cryptic comments about what has been called "information bias," arguing that such bias may be an illusion. We wrote: "information that changes the structure of the market should (arguable) change the circumstantial choices made therein." This argument piqued the curiosity of Cummings et al., who devoted several pages to wondering what we could have meant. The economic-theoretic analyses that I have discussed above provide a sound basis for further explicating our argument.

Stripped to its barest essentials a contingent market offers a public policy for approval or disapproval. From the respondent's perspective any such policy is a pairing of commodities delivered and payments exacted. Thus, the rational respondent bases her contingent market decision on (1) the value to her of the commodity or amenity offered, (2) the rule by which the agency decides whether or not to provide the commodity, and (3) the rule that determines the payment exacted from the respondent. Note that all three are relevant to policy evaluation and a change in any one of them could chnge CVM results. However, only item (1) directly enters the standard economic model for valuing nonrival goods. In this vein, the concept of incentive-compatibility addresses the issue: do (2) and (3) encourage reporting of (1) inconsistent with the standard economic model of value?

The empirical evidence that Cronin and Berzeg, and Rowe et al., inter alia, have marshalled to support charges of "information bias" shows that changes in (1), (2) and/or (3) tend sometimes to change reported WTP. We emphasize that contingent policy evaluations should be expected to change as these things change. A policy evaluation tool with results invariant to important changes in these conditions would surely be misleading and uninformative. Exit "information bias."

Nevertheless, for economic valuation of nonrival goods, the issues of incentive compatibility and the satisfactoriness of PPI indicators remain. As Hoehn and I have shown, careful analysis of the CVM structure with respect to (2) and (3) serves to identify structures that generate satisfactory data for nonrival goods valuation.

Note that markets can be viewed as a special case of a more general class of resource allocation mechanisms or policy choice mechanism, all based on individual utility maximization within the constraints imposed by fully specified public decision rules (item 2, above) and individual payment rules (item 3). It seems logical to expect that satisfactory contingent valuation designs could be constructed for any member of this class of mechanisms. Especially when the commodities to be evaluated are both nonrival and nonexclusive, contingent valuation formats may fruitfully be designed consistent with the more general class of policy choice mechanisms. again, the policy choice referendum format is clearly admissable (and is a member of the same class of resource allocation mechanisms that includes traditional contingent markets).

C. CONCLUDING COMMENTS

The economic-theoretic approach has been fruitful in clarifying the incentives facing a CVM respondent. A class of satisfactory BC estimators has been identified. Some empirical results once thought anomalous -- including but not limited to those pertaining to so-called "information bias" and the divergence between WTP and WTA -- are now seen as rational an predictable responses to the costs and opportunities inherent in contingent markets. Some simple principles have emerged that will be useful in improving CVM by reducing the extent of benefit understatement associated with compensating WTP and the prevalence of results that seem anomalous.

But perhaps most important, our work leads us to be conscious that contingent markets are not devoid of incentives for reasoned decision making therein. Further, there exists a class of contingent valuation mechanisms that are immune to strategic manipulation. Together, these findings place CVM in a new perspective.

Simplistic dismissals of CVM -- "it is utterly devoid of incentives for reasoned decision making," and "it is riddled with opportunities for strategic behavior" -- must themselves be dismissed. Arguments that practitioners must consciously downplay any association between CVM results and policy outcomes, in order to contain "strategic bias," must be rejected. on the contrary, policy relevance would appear to enhance the incentives for careful value formulation. A dilemma commonly claimed to bedevil CVM -- "increased policy relevance causes strategic bias, while decreased policy relevance causes hypothetical bias" -- simply does not exist, if one uses CV mechanisms selected from the class of satisfactory BC indicators.

The defense of CVM no longer rests on empirical case study evidence that seems to fly in the fact of reason. We have shown that theoretical analysis of the incentives inherent in CVM offers some support for the method, as well as some suggestions for its improvement.

ENDNOTES

CHAPTER VIII

*) My experience with the contingent valuation method was gained in the course of research sponsored by U.S. Environmental Protection Agency, U.S. Fish and Wildlife Service, Resources for the Future, Inc., and New Mexico Agricultural Experiment Station. The viewpoint expressed in this paper has been influenced by more than a decade of interaction with may of my colleagues in environmental economics. My close working relationship for the last several years with John P. Hoehn has provided countless opportunities to develop and refine the argument.

1) Some economists tend instinctively to question whether citizens would rationally behave as though they could expect to have any influence on policy. Their skepticism is apparently based on the standard free-rider model, itself a result of single-period analysis of voluntary provision of pure public goods. However, recent theoretical models of repeated games with uncertain ending periods have demonstrated that free-riding is not always individually optimal. It may be rational to cooperate in maintaining the institutions of social stability.

Empirical evidence from elections indicates that many people participate and that, within the limits implied by the electoral system, they pursue their self-interest therein. Savers, investors and those who favor limits to redistribution tend to vote for Republican and/or "conservative" candidates. Debtors, low-wage earners and welfare recipients tend to vote for Democratic and/or "liberal" candidates. The "misery index," which rises with unemployment and inflation, remains the best predictor of election results: high levels of this index bode ill for incumbents.

One need merely appeal to casual observation to confirm the considerable investment of time and effort expended by ordinary individuals in gathering and processing political and policy-related information and attemping to influence policy via individual and voluntary group activities.

2) If equivalent measures of value are sought, the results of formulation error are not so clear. There are two problems to solve: (i) the "with policy," or subsequent utility level must be found by maximizing utility given the subsequent opportunity set, and (ii) expenditure must be minimized subject to utility constrained at the subsequent level. Formulation error at stage (i) would understate subsequent utility and thus expenditure, while error at stage (ii) would overstate expenditure. The final outcome is ambiguous when equivalent measures of value are used.

3) Thus, the now commonplace empirical finding -- that CVM tends to generate larger differences between WTP and WTA than Willig (1976)

and Randall and Stoll (1980) would predict -- is in now way inconsistent with the satisfactoriness of CVM in the compensating mode.

4) This is a clear departure from my previous position on this issue (Brookshire et al., 1980).

IX. DOES CONTINGENT VALUATION WORK?

BY
Richard C. Bishop
Department of Agricultural Economics,
University of Wisconsin
and
Thomas A. Heberlein
Department of Rural Sociology
University of Wisconsin

Two tasks have been assigned to us: First, we have been asked to critique Part I of this volume, the part prepared by Cummings, Brookshire and Schulze. Following the precedent they establish, we will refer to Part I as CBS. Second, we are to give our own assessment "of the promise, strengths, and weaknesses of CVM."

TO accomplish these assignments we must begin with some background material that will help to justify our views. As CBS point out, our team at Wisconsin is investigating the validity of CVM by comparing contingent values for hunting permits with values from actual cash transactions. Our experiment involving goose hunting permits has been described by CBS, but a brief discussion will help clarify some additional points. More importantly, we will introduce some preliminary results from a new experiment involving deer hunting permits. Results here are germane to a number of questons raised by CBS as well as to our own views about the accuracy of CVM.

Drawing on these experimental results, the second section will comment on CBS. Let it be said at the outset that we find much to commend in their work. It is certainly timely to systematically assess what has been learned and to chart a course for the future. Their stubborn insistence on clearly stating and testing hypotheses is laudable. It is also high time that researchers explicitly recognize the potential pitfalls of using market data in TCM and HPM. Surely CVM will work better in some contexts than others. Hopefully, a systematic, empirically verified, set of conditions for successful application of CVM will be developed. On these and many other issues, we heartily support CBS in their efforts to evaluate what has been learned during 20 years of research on CVM. However, we find much that is questionable in the specifics of CBS's presentation. They are not very definite when drawing conclusions concerning bias. Their endorsement of iterative bidding is not well founded empirically. They need to recognize the potential usefulness of field experiments as powerful complements to laboratory experiments. These and other points will be raised and clarified in section B.

In our own assessment of CVM -- to be presented in section C --- we will attempt to answer the question posed in our title: Does contingent valuation work? Obviously, to state the issue in this way is an oversimplification. There is not a categorical answer, Rather, the question is really: How well does CVM work?

Our position on CVM is interesting in light of where we started. In 1978, when we first began our own research on CVM, we were among the most cynical. It would not have surprised us to learn that CVM produces totally meaningless results, In the coming pages we will argue that, while CVM is

inaccurate even under the best of circumstances, it is still capable of producing policy-relevant values when competently applied in suitable situations.

A. EXPERIMENTAL RESULTS

1. Goose Study Design

Since this study has already been published (Bishop and Heberlein, 1979; Bishop, Heberlein and Kealy, 1983) and summarized by CBS, we will be brief, but some clarification is desirable. Our three samples of hunters had been issued permits to take one Canada goose in the Horicon Zone, an area of 24,600 acres in east central Wisconsin where geese concentrate each fall. These permits applied only to the period between October 1 and October 15, 1978, and a hunter was allowed only one Horicon Zone Permit for the entire 1978 hunting season. The permits were free. A total of almost 14,000 permits was issued.

The first sample (237 hunters) received actual cash offers by mail to forego their 1978 Horicon hunting opportunities. Dollar amounts were assigned at random between $1 and $200. The second sample consisted of 353 people who were involved in a mail survey in which the principal CVM question was worded identically to the actual cash offers, except that the hypothetical nature of the proposed transaction was emphasized. Other CVM questions including WTP questions were also included. The third sample (300 hunters) was surveyed after the goose hunt and the results used to estimate a TCM model. All samples were surveyed for specific attitudes regarding goose hunting and general socioeconomic characteristics. In all cases response rates exceeded 80 percent.

2. Goose Study Analysis

The responses to the actual cash offers were either yes or no. These dichotomous responses were analyzed in a logit model of the form

$$\pi = (1 + e^{\beta Y})^{-1}$$

where π is the probability that a hunter will accept an offer, Y is a vector of explanatory variables, and β is a vector of coefficients. Some results from the maximum likelihood estimation of this model for the actual cash offers and parallel contingent market are given in Table 9.1. The explanatory variable, ln Dollars, is simply the natural logarithm of the dollar offer amount. Model 2 includes a second explanatory variable, Commitment, which is a four item attitude scale expressing the level of commitment each hunter had to goose hunting with larger values expressing greater commitment. Both explanatory variables have the expected signs, i.e., larger dollar amounts would be expected to increase the probability of selling while increased commitment would be expected to reduce the probability of selling. Chi-squared tests comparing actual cash equations with respective CVM equations showed statistically significant differences at the .05 level for both models.

Examining the coefficients in Model 1 indicated that increasing the dollar amount had a much stronger effect for the actual cash offers than for the hypothetical offers. Thus, when the expected value of a permit was calculated, it was $63 for actual cash offers and $101 for the hypothetical ones. To obtain these values we truncated the model at $200, the largest amount for which we had data. The medians, defined as the dollar amounts where the probability of acceptance was 0.5, were $29 and $80 for the cash and hypothetical markets, respectively. A parallel willingness-to-pay

question was asked where respondents were requested to assume that they had not received a permit and asked whether they would pay a specified amount, again set randomly between $1 and $200. The expected value here was $21. The median was $5.

Table 9.1

Regression Analysis of Simulated and Contingent Markets for Willingness-to-Sell Goose Hunting Permits a

	Model 1		Model 2	
Explanatory Variables	Simulated	Contingent	Simulated	Contingent
Constant	3.99** (.66)	3.24** (.54)	1.72 (.98)	-.58 (.81)
ln Dollars	-1.18** (.18)	-.74** (-13)	-1.16** (.18)	-.84** (.14)
Commitment			.21** (.07)	.40** (.07)
N	189	306	189	306

a Standard errors are given in parentheses.
** Indicates coefficient significantly different than zero at .01 level.

To set the record straight, it needs to be stated that Bishop and Heberlein (1979) emphasized that all results in that paper were preliminary, including the TCM values which were reported by CBS to be between $11 and $45, depending on assumed value of time. Later modifications of our TCM model, which we believe are more: in keeping with the current state of the art for travel cost work, yielded a value of $32 assuming a value of time equal to 50% of the income rate. This is the value which we prefer to use as the TCM result for our study. See Bishop, Heberlein, and Kealy (1983) for further discussion.

3. Goose Study Interpretation

How are such large differences in values (ranging from $21 to $101) to be explained? Setting aside for the time being the travel cost result, what about the apparent errors in CVM values for willingness-to-pay and willingness-to-accept-compensation?

Let us explicitly state that our actual cash transactions are not a perfect criterion against which to evaluate CVM results. As CBS repeatedly emphasize, we would all be quite satisfied if CVM approximated values from a real market. Our cash transactions do not fully measure up to this standard. Disequilibrium may be a factor. Respondents to our actual cash offers get only one opportunity to engage in a transaction while real markets, even for durables such as automobiles and houses, generally involve repeated transactions over long periods of time. The opportunities to gain

experience, obtain information, and "research preferences" must be much more extensive in real markets. To go a step further, our cash offers may well share some of the bias problems that CBS have outlined for CVM. To take an extreme view one might even speculate, for example, about strategic bias. Suppose that individuals receive a cash offer from us as part of a single experiment and that they see some advantage in influencing final results in an upward direction. They might well refuse offers which they would accept in a real market in order to further their long run goals. We have repeatedly called attention to the fact that our cash offers do not constitute a full-blown market by referring to our approach as a simulated market.

Still, the simple result remains that people did not respond to hypothetical offers in the same way that they responded to cash offers. Our results clearly show that people refused hypothetical amounts that they would have accepted in actual dollars. Why? As CBS points out, we attribute this behavior to the artificiality of CVM procedures (Bishop, Heberlein, and Kealy, 1983). Look at Table 9.1 again, this time focusing on Model 2. While In Dollars had a stronger influence in the simulated market, commitment had a much stronger influence in the contingent market. Our interpretation is that people have never tried to value goose hunting before and do not know what they would accept when confronted with a questionnaire. To answer, they fell back on their commitment to goose hunting and related tastes and preferences more than they would have if real money was before them. Real money draws more attention than hypothetical money and helps them to "research their preferences" in a more realistic economic milieu. There is more incentive to consider a real offer because the losses from making an error are greater. As we have said before, money is a strong stimulus and real money is a stronger stimulus than hypothetical money. This argument clearly parallels CBS's treatment of bias due to hypothetical payment.

Like most researchers, we have not been able to resist the temptation to reach beyond our empirical results and speculate about their broader implications. Suppose we are correct that hypothetical bias in the form just described is the central problem in CVM. In which direction does the bias lie? Clearly the results presented here indicate that CVM willingness-to-accept-compensation will be an overestimate. To move to the willingness-to-pay side is more tenuous because we had no actual cash transactions involving payment for permits. Nevertheless, we argued (Bishop, Heberlein and Kealy, 1983) that people respond to the artificiality of CVM by giving conservative responses. They refuse hypothetical offers unless they are certain they really would accept. If this same conservatism is exercised on the willingness-to-pay side, people will indicate refusal to pay unless they are relatively certain that they really would pay. This would make CVM willingness-to-pay an underestimate. This appears to be consistent with the empirical evidence we have. First, attempts to work income into various logit and travel cost equations consistently produced coefficients that were small and insignificant. This absence of an income effect appears to imply that willingness-to-pay real dollars should be $63 as well, except for the possible influences of disequilibrium mentioned above. Second, we did have a measure of actual willingness-to-pay in the TCM estimate of $32. By comparison, our CVM value using the hypothetical offer to sell permits to hutnters at fixed prices was $21. The CVM survey also included an open-ended question asking the respondent to write in maximum WTP. Here,

the mean was $11 (Bishop, Heberlein and Kealy, 1983, p. 627). This was our viewpoint when we initiated the deer hunting study. Empirical research can hold surprises, as we shall see momentarily. First, however, CBS makes rather prominent mention of the unpublished criticisms of our goose study analysis by Carson and Mitchell (1984). Let us digress, therefore, to address their concerns.

Carson and Mitchell (hereafter CM) claim that two groups of hunters included in our analysis should be eliminated because their responses are invalid, or as CM put it, they were not genuinely participating in the studies. They show that when these "nonparticipants" are eliminated from the analysis, the estimated values provided by the CVM and the simulated market are statistically the same. We disagree with the assumptions underlying their reanalysis, and argue that our original estimates are correct.

First, in the cash market only, CM eliminate the 15 hunters who neither cashed the check nor returned it, refusing the offer. We classified these hunters as refusals to sell, while CM claim they are nonparticipants. Since each hunter had already received his/her permit, and since the permit would not be invalidated unless they cashed the check, it is highly likely that most of these "nonparticipants" simply took the easy way of refusing, that is, destroying the check. Further, Hanemann's (1983) analysis found no effects of a nonresponse dummy variable on the estimated cumulative density function for acceptance.

The second group of "nonparticipants" eliminated from CM's analysis are a proportion of those who refused to sell at amounts above a particular truncation point. They specify the appropriate point as that "beyond which no further sustained (statistically) significant increase in the acceptance rate occurs." Therefore, they eliminated from the cash market analysis those respondents who refused to sell at $75, $100, $150 and $200 (i.e., ten percent of the total) and from the hypothetical market analysis those who refuse to sell at $50 and above (over 50 percent of respondents!). They suggest that these respondents are not genuinely participating in the study, but are "protesting" the study or the idea of selling goose permits in an open market by refusing to sell at a price well above their true permit value.

On the face of it, we find it implausible that many hunters would forego $75, $100, $150, or $200 for the privilege of expressing such an opinion unless the goose permit itself were very close to the amount offered (and refused). The fact that the refusal rate levels off between $50 and $200 simply indicates that most of the people who did not sell for $75 would not sell for $100 to $200 either. These hunters are those who place a high value on opportunities to hunt at Horicon, and it would take perhaps much more than $200 to buy their permits. CM's analysis assumes that this minority group of high-value hunters does not exist, and/or that their values should not be included in an estimate of "public" values. Had we been able to offer larger amounts, $500, $1000 and so on, we might have found the point at which the last of these high-value hunters would give up the permit, but it would certainly be greater than $200, and our estimate of $63 is therefore a conservative one (as noted by Hanneman).

Detailed analyses of several attitudinal variables provide further support for our hypothesis and refute CM’s hypothesis of "protest." Attitudes toward valuation research and attitudes toward paying for hunting privileges were <u>not</u> related to WTA, when the dollar amount of the offer

was contolled. Further, hunting commitment did have a direct effect on refusal to sell, controlling for dollar amount, in both the simulated and cash markets.

In sum, we disagree with CBS's statement that "Carson and Mitchell demonstrate, using Bishop and Heberlein data, the lack of significant difference between hypothetical and 'actual' payments" (p. 108); however, we will await the publication of Carson and Mitchell's comment to make a more comprehensive response.

4. Deer Study Design.

Our reasons for developing a second simulated market experiment extended far beyond mere replication of the goose study results. The goose study did not include simulated market evidence on willingness-to-pay, yet researchers have been more interested in willingness-to-pay measures than in measures of willingness-to-accept. Our valuation mechanism in the goose study (take-it-or-leave-it offers) was rather unorthodox. Most past CVM studies have used bidding games or open-ended valuation questions. As CBS point out, many researchers prefer bidding games because they feel that the bidding process encourages more carefully reasoned consideration of respondents' maximum values. With respect to the goose study, one has to wonder how bidding would have affected both the simulated market and CVM results In a broader perspective, we also wanted to determine whether the large differences between WTP and WTA, documented consistently in CVM studies, carry over to treatments involving actual cash transactions.

To address these issues, we conducted a study of the value of deer hunting at Sandhill Wildlife Demonstration Area in Wood County, Wisconsin. This is a 12-square mile wildlife research area with a deer-proof fence around the perimeter. Recent research on deer has emphasized management for trophy bucks. In order to maintain the deer population within habitat limits and satisfy multiple-use goals for the area, a deer hunt has been permitted over the past several years. During the past three years, hunters were allowed to take one deer of either sex using their regular Wisconsin deer hunting license, In addition to that license, each Sandhill hunter had to be the winner of a lottery. For the 1983 hunt, which took place on November 12, 150 permits were issued for almost 6,000 applications.

For purposes of the -experiment, the 150 successful applicants (i.e., lottery winners) were divided into two groups of equal size. The first group was told that we intended to purchase four Sandhill permits from those who bid the lowest amounts in a sealed-bid auction. Each successful bidder would be paid his/her bid and would not be able to hunt at Sandhill in 1983. The other group of 75 successful applicants received contingent valuation surveys with parallel wording.

A random sample of 600 individuals was drawn from the pool of unsuccessful applicants. Half of these individuals were involved in actual auctions to buy a total of four Sandhill hunting permits issued by the state for research purposes. The other half were involved in comparably worded contingent valuation auctions. Four different auction systems were used. One-fourth of the participants were given the opportunity to simply submit a sealed bid. Their initial bid was the bid that was entered into the auction. A second auction which we will term Bidding Game 1, involved an initial sealed bid. However, these individuals were later contacted by telephone and allowed to raise or lower their bids in a bidding game format. The third auction mechanism involved an initial contact by mail which

included a fixed initial bid chosen at random between $1 and $500. Respondents could respond positively or negatively to this initial bid and it served as the starting point for bidding games conducted during later telephone interviews. This will be designated as Bidding Game 2. The fourth auction mechanism involved sealed bids. However, in this case respondents were assured that if their bid was of the four highest bids, they would not be required to pay their full bid, but a lesser amount equal to the fifth highest bid across all the auctions. Thus, this treatment was like the Vickery auction discussed by CBS, except that it was a "fifth price" auction rather than a second price auction. The economic incentives are the same as in the Vickery format, with expected utility theory indicating that hunters would bid their full compensating surplus in such a situation. All study subjects were surveyed by mail after the bidding was completed and all were paid $5 for timely participation, including return of the questionnaire.

5. Preliminary Results.

A total of 683 hunters (91%) participated fully in the auction. Actual cash bids to accept compensation ranged from $25 to $1,000,000. The $1,000,000 bid was interpreted as a response of "not for sale" and deleted from the analysis that follows. The next highest bid was $20,000. Accepted bids were $25, $62, and two bids of $72. Hypothetical bids to accept compensation ranged from $0 to $20,001. WTP cash bids to buy a permit ranged from $0 to $200, with accepted bids being $200, $177, $152, and $150. Only the $152 bid came from the Fifth Price Auction and this person actually paid $142.

Considerable further analysis remains to be done on the results of this experiment. Bid functions have not been estimated for example, so we Can not yet say whether commitment to hunting, income, and other variables played a systematic role in determining bids. Our TCM work is only now getting underway. Still, the preliminary results do suggest some tentative conclusions.

Table 9.2 shows means and other statistics for the willingness-to-accept-compensation side of the experiment. The mean cash offer of $1,184 was not significantly different from the CVM mean bid of $833. The estimated standard deviation of the bids was quite large.

Table 9.2

Willingness to Accept Compensation For
Sandhill Deer Hunting Permits.

	Mean	Median	Mode	Standard Deviation	N
Cash Offers a	1184 *	550	1000	2475	70
Hypothetical Offers	833 *	102	100	2755	70

a $1,000,000 cash bid excluded as an outlier.

* Indicates that mean of cash offers and mean of hypothetical offers not statistically significant at the .05 level.

For willingness-to-pay, our preliminary results are given in Table 9.3. Cash offers averaged between $19 and $25 in the different auction formats and the null hypothesis that these means were equal could not be rejected at the .10 level. Mean hypothetical bids varied between $31 and $44 and there were also no significant differences among the auction formats. Comparisons of cash and hypothetical bids within auction formats shows that the hypothetical bids are significantly different at the .10 level in three out of the four cases. In all four cases the mean hypothetical bids were larger.

Next consider the effects of bidding. The format designated as Bidding Game 2 in Table 9.3 most closely parallels the traditional CVM bidding game. Respondents here, it will be recalled, answered yes or no by mail to a starting bid. Then, bidding by telephone followed using the starting bid at the outset. Table 9.3 reports the mean final bids, which are amazingly close to those from the other treatments. The telephone bidding process did not produce significantly higher or lower results than the Sealed Bid Auction, the Fifth-Price Auction, or Bidding Game 1. This was true whether the comparison was across hypothetical or cash auctions.

Table 9.3

Willingness To Pay For Sandhill Deer Hunting Permits.

Auction Format	Mean	Median	Mode	Standard Deviation	No. of Observations
Sealed Bids:					
Cash	$24	$15	$ 5	$35	68
Hypothetical	$32	$11	$10	$64	71
Bidding Game 1 [a]					
Cash	$19 *	$10	$ 5	$23	65
Hypothetical	$43	$21	$ 0	$58	62
Bidding Game 2 [b]					
Cash	$24	$15	$ 0	$30	68
Hypothetical	$43	$20	$ 0	$69	69
Fifth Price					
Cash	$25 *	$20	$10	$30	69
Hypothetical	$42	$21	$10	$70	70

a Respondents set initial bids.

b Initial bids chosen at random.

* Indicates hypothesis that mean cash bid equaled mean hypothetical bid for these auction formats was rejected at the .10 level.

In Bidding Game 1, the respondents were asked to submit sealed bids by mail. If they read the "fine print" carefully, they would have seen that the possibility of later changing the bid was kept open, but this possibility was not emphasized in order to get valid sealed bids, yet make the contracts for cash offers legally binding. No mention was made of later telephone bidding or any other mechanism for changing the bids. The initial bids averaged $14 and $25 for the cash and hypothetical groups respectively. Telephone bidding caused 42% of the cash bids to increase. The final bids averaged across the entire subsample increased by $5 to reach the $19 final bid reported in Table 9.3. For the hypothetical sample, the mean final bid was $43, an increase of $18. Of the 62 people we were able to recontact, 52% increased their bids. Comparing the mean increases showed that people tended to increase their bids more in the contingent auction than in the actual cash auction, with the difference being significant at the .01 level.

By way of summary, preliminary results from the deer experiment seem to point to four conclusions:

1) The large differences between WTP and WTA compensation so often observed in CVM studies carry over to transactions involving real money and real recreational opportunities. In our contingent auctions, WTP averaged $40 across all auction formats combined, while WTA averaged $833. When real money and real permits were involved the difference was slightly larger at $23 versus $1,184. This latter result is consistent with findings of Knetsch and Sinden (forthcoming). Large differences between WTP and WTA are not simply a phenomenon of CVM.

2) WTP was significantly higher in the contingent auctions than in the cash markets. We suspected a tendency to bid higher in the cash auction measure of WTA, but the difference was not statistically significant in this data set. We will return to this point momentarily.

3) Bidding did not seem to make much difference. People in Bidding Came 1 did tend to raise their offer amounts and the tendency was stronger for the hypothetical bids. Those tendencies, however, did not produce changes that were large relative to variations in mean bids due to intersample differences. Bidding Came 2, which closely parallels traditional bidding games used in CVM studies, produced results nearly identical to other auction formats.

4) As one might expect, based on the literature on experimental auctions cited by CBS, the Fifth-Price Auction did not produce the significantly larger bids that theory would lead one to expect. Vickrey auctions seem to be of questionable value in CVM studies, a point that we will discuss further in our evaluation of CBS in Section II below.

6. Deer Study Interpretation And Plan For Further Research.

These results contradict what we expected based on the goose study. As noted above, we expected CVM estimates of WTA to be much larger than cash experiment estimates. If anything, the WTA results tend in the opposite direction and the difference is not statistically significant. While our evidence was not as strong, we thought that a good case existed for arguing that CVM estimates of WTP tend to be underestimates of true WTP. The deer study had hypothetical WTP offers significantly higher than comparable results based on cash offers. How are these differences to be explained?

Of course, a larger number of hypotheses could be stated to try to explain these differences. As our analysis continues, and particularly as we estimate bid functions, additional possible explanations may become apparent. At this writing, our best guess is that a large part of the difference between the goose study and deer study results are attributable to the added uncertainty present in the deer study.

The goose study respondents made their decisions under relative certainty. If they accepted our fixed, predetermined offer they received the amount of money offered. If they rejected the offer, they maintained the opportunity to hunt a goose.

The problem for our deer hunters was more complicated. The effect Of bidding on the cash position and hunting opportunities of any given respondent depended on how much she or he bid and the bids of all other auction participants. The bidding behavior of others, particularly given the absence of any information from past auctions, must have been very uncertain. As CBS point out in some detail, people do not seem to react to uncertainty in ways that are consistent with what utility theory would lead us to expect. Theory would lead us to expect very similar behavior in simulated markets involving fixed take-it-or-leave-it offers and simulated

markets involving various bidding frameworks, particularly the Fifth-Price Auction. However, respondents have reacted to the added uncertainty inherent in bidding against others in ways that led to very different results. We suspect that people tended to adopt a "heuristic" which led them to behave very conservatively in response to the uncertainty about other's bids.

Consider the cash auction where we offered to buy four permits from the lowest bidders. Participants in this auction had won the lottery with odds of 0.025:1 (150 winners out of 6,000 applicants). They were then asked to state the minimum amounts they would accept. People may have figured that by stating a high bid they increased the risk of losing the auction, but then they could always go hunting. If a high bid was stated, but other bids were even higher, the bidder would lose the hunting opportunity but receive a relatively large amount of money. Making a relatively low bid improved the chances of winning, but winning the auction would entail loss of the hunting opportunity and the monetary gain would be small. We suspect that this sort of logic tended to lead our study participants to state relatively large bids in the cash auction to estimate WTA, bids in excess of their true compensating surplus. The same rationale could have been active in the CVM treatment, but naturally would have been less powerful because study subjects knew that they could go hunting regardless of how they responded.

On the WTP side, this same conservatism would work in reverse. Consider the point of view of a hunter drawn to participate in the cash auction. If he or she bid a relatively low amount, then the result would probably be loss of the auction, but there was some chance that others would bid even lower amounts, thus making the person in question a winner. In this way, our auction provided a small chance of a real bargain for those who bid relatively small amounts. Certainly bidding higher would improve the chances of winning, but more of the potential compensating surplus will be lost due to the higher cash payment required. People may have had a tendency to bid toward the lower side of their compensating surplus. We hoped that the Fifth-Price Auction would reduce this tendency, but apparently uncertainty was the overriding consideration. In the CVM auctions, hunters knew that they would not have the opportunity to hunt regardless of their answers and tended to react by bidding higher than they would have in the cash auction.

This scenario, though plausible, is only speculative at this stage. More definite conclusions must await further research. The 1983 Sandhill study involved only four permits because of legal constraints that are no longer binding. For the 1984 hunt, we can deal in any number of permits so long as the requisite number of hunters to meet biological objectives is present. This will make it possible to construct a 1984 study like the goose study. Simulated market participants on both the WTP and the WTA sides will receive opportunities to buy and sell, respectively, permits at predetermined prices. This will make uncertainty about other bids irrelevant. Our guess is that CVM WTP will tend to increase slightly and that CVM WTA compensation my fall a bit. More importantly, we hypothesize that this new format will have a large upward effect on simulated market WTP and a large downward effect on simulated market WTA. Using SM to symbolize "simulated market" our hypothesis is that:

$$\text{CVM WTP} < \text{SM WTP} < \text{SM WTA} < \text{CVM WTA}.$$

However, we expect large differences between SM WTP and SM WTA to remain.

Having thus stated what we have learned about CVM from our own research let us return to CBS for some implications.

B. IMPLICATIONS FOR CBS.

CBS have provided a great deal of food for thought in the first part of this book. Much is said to which we can readily agree. Rather than dwell on these points of agreement, we will focus in on areas where we disagree or at least think more should be said. The evaluation of CBS will be organized around a series of rhetorical questions in the hope of focusing attention on major issues.

1. Is Contingent Valuation Biased?

CBS make many good points in this regard, but even after reading them carefully, we are not quite sure where they stand on the question of bias.

What is meant by bias here? CBS suggest (p. 13) that "market prices are appropriate measures of the 'benefits' (social welfare) of concern in cost-benefit assessments and, therefore, represent a standard for accuracy, or 'appropriateness', against which CVM measures are to be compared". While we will raise some questions later regarding specific interpretations of this statement in the context of WTA, the basic principle is clear: CVM values are accurate to the extent that they approximate values that would obtain in a well-functioning market.

This is why we believe that our experimental results are powerful. Although -- as noted above -- our simulated markets for hunting permits lack some of the characteristics of real markets, they should provide considerable information about how comparisons with real markets would come out. Furthermore, the comparisons are being conducted under rather ideal conditions. Hunting permits are not a public good, since the excludability condition for private goods is fulfilled. Furthermore, the commodities -- goose hunting or deer hunting opportunities -- are well-known to the study subjects. Vehicle bias should not be a problem, since both the hypothetical and cash transactions employed the same vehicles. All study subjects, whether in the real or hypothetical markets, had the same information. The only difference in the treatments was that part of the transactions involved hypothetical payments and recreational opportunities and part invovled real payments. Clearly if contingent valuation is capable of giving unbiased estimates of real values, it should have done so here.

The results, however, indicate bias. People were more willing to sell their goose hunting permits for real dollars than they indicated they would be in the contingent market. Preliminary results from the deer study indicate that in an auction framework, CVM will overestimate willingness-to-pay. On the WTA side of the deer hunting auction, bids varied too widely to say for sure, but it appears that CVM may have erred slightly on the low side.

How would this bias be classified within the system described by CBS? Hypothetical bias related to the lack of real transactions appears to be the problem. As we have said before, money is a powerful stimulus and real money is more powerful than hypothetical money. In fairness to CBS, they seem to be very explicit in recognizing this point. For example, citing us and other studies, they point out (p. 107) that, "actual vs. hypothetical payment does result in different choices" (emphasis in original). However, somehow this does not seem to be a major point in their overall argument. In an earlier section of Part I (p. 29), they refer to hypothetical bias as "one of the most important unresolved issues for any assessment of the efficacy of CVM". In their Executive Summary, they

mention our result, but quickly point out that Carson and Mitchell cast doubt on the conclusions. Similar, though less specific, questions are raised about Bohm's findings and those of Slovic and others. If one read only the Executive Summary, one would come away with the impression that the jury is still out on this question. With the added evidence from the deer study -- to which, admittedly, CBS did not have access, since it is as yet unpublished -- we think the evidence for bias related to hypothetical payment is rather convincing.

Furthermore, this source of bias lies at the crux of the matter. CVM'S dominant characteristic is the hypothetical character of the transactions. Starting point bias, information bias, vehicle bias, and biases relating to perceptions and framing may well arise in circumstances that are less ideal than ours. However, even if these problems are solved to a satisfactory approximation, bias due to hypothetical payment will still be a threat. Stated differently, no matter how closely the "Reference Operating Conditions" (ROC) proposed by CBS in Chapter VI are met, hypothetical bias will remain. In fact, it is hard to imagine any real world setting where the ROC's would be more closely met than in our experiments, except that we measured only WTA in the simulated market for goose hunting permits. Hypothetical bias deserves more explicit recognition by CBS outside of Chapter V.

2. Do CBS Deal Adequately With Accuracy Issues ?

To ask "What Is Accuracy?" in the context of nonmarket evaluation is long overdue. Thus, CBS have produced much that is thought-provoking and we hope that they and others will pursue this topic with diligence. However, we have some serious reservations about the specifics of their accuracy assessment. It may be necessary to accept accuracy no better than ±50 percent in estimates from CVM, TCM, HPM, and market data, but CBS's arguments for such a limit are hardly convincing.

CBS claim (p. 190) that, "The range of possible error for the CVM derived solely from potential biases is easy to establish". They then cite Rowe et al. (1980) as showing that the sum of starting point, vehicle and information bias can be as large as 40 percent. They also cite Schulze et al. (1983) as showing that payment cards may produce results as much as 40 percent lower than iterative bidding. Applying these percentages leads CBS to conclude (p. 191) that "CVM is not likely to be more accurate than ±50 percent of the measured value" (emphasis in original).

Surely such a wide range of error need not be accepted. Rowe et al. are not using the term 'bias' in its strict sense of deviations from the "true" value. Instead, they showed that varying starting point, vehicle, and information can cause final bids to vary greatly. CBS argue (and we add our support below) that experimental techniques should be very helpful in reducing such variation by indicating which CVM techniques come closest to approximating true values. Surely many of the sources of error found by Rowe et al. can be reduced or eliminated through experimental studies. As for the results of iterative bidding found in Schulze et al. (1983), either iterative bidding helps bring people closer to their true values or it does not. Experimentation should be able to produce strong evidence one way or the other. Thus, the studies cited by CBS are not indicative of the magnitude of errors that are inevitably present in CVM and that must be accepted in setting error bounds.

Similar problems may exist in CBS's assessment of the errors in value

estimates from market information. Unfortunately, the paper they draw on (Coursey and Nyquist, 1983) is unpublished and therefore unavailable to us. We were unable to follow the argument as described by CBS.

Thus, we would question whether CBS' assessment of bounds for CVM estimates and market demand analyses are meaningful. This is not to say that the bounds are necessarily less than or greater than 50 percent. More research is needed to implement the specifics of CBS's sound overall ideas about accuracy.

Furthermore, an important concept is missing from CBS's exposition on scientific accuracy. This is the concept of "calibration". When a new method of scientific measurement is developed it is often necessary to calibrate it against old methods. It may prove feasible through experimental studies to calibrate CVM methods that can then be used in the field to arrive at more accurate values. Thus, establishing error bounds on existing CVM techniques is a worthwhile goal, but reducing those bounds through calibration should be the long-run goal.

3. Does Iterative Bidding Improve Accuracy?

CBS give a rather strong endorsement to iterative bidding. They repeatedly emphasize that this procedure emulates "market-like" processes, helping respondents to "research their preferences". Also, the experimental literature is cited (see, for example, p. 83) to show that in auctions people may require several rounds of bidding before they learn their optimal strategies. Iterative bidding allegedly provides a substitute for this learning process.

Considerable evidence can be mustered to the contrary. The evidence is not strong enough to reach categorical conclusions yet, but there are substantial indications that iterative bidding biases CVM results.

CBS (pp. 59-66) review a great many studies that have attempted to test for starting point bias in traditional bidding games. Some found an upward bias, while several others did not. We would submit that all of these studies provided relatively weak evidence because they involved only two, or at most three, starting bids. Furthermore, sometimes the range of Starting bids was too small to pick up starting point bias.

To further examine the question, members of our research team have recently analyzed data from three studies employing bidding games. These include a CVM study of the value of scenic beauty along the Lower Wisconsin River (see also Boyle and Bishop, 1984); the deer hunting permit study, Bidding Game 2 as reported above, and a study of the value of sport diving around offshore petroleum structures (Thompson and Roberts, 1983). In the first two studies starting bids were randomized across a range of values that were deemed ex ante to be plausible. In the Thompson and Roberts study, six alternative starting bids were used ranging from $20 to $400 for a year of diving.

To test for starting point bias, we hypothesized a linear relationship of the following form:

$$BF = a + b\ BS + e$$

where BF is the final bid, BS is the starting bid, a and b are constants and e is a random disturbance term. The equation was estimated for four different sources of data: (1) the Wisconsin river contingent valuation results; (2) the deer study results from contingent bidding; (3) the

deer study results from cash bidding, and (4) the Thompson and Roberts study. The results are reported in Boyle, Bishop and Welsh (1984). The estimate of b was positive and significantly different from zero at the 0.01 level for all three CVM data sets. The estimate for b was negative and not significantly different from zero at the 0.10 level for the cash bidding for deer permits. We would interpret this as evidence for the hypothesis that the starting bid has a significant positive influence on final bids in contingent markets. Furthermore, this phenomenon does not seem to be present once real money becomes involved.

By way of a caveat, we should say that these results are new. Discussions are already underway with Alan Randall about their validity. Randall would argue that perhaps our range of starting bids included some that were too far removed from most people's final bids. He suspects that when the bidding process starts at such high levels people tend to become tired of and bored with the bidding process. They then terminate the bidding by accepting bids which are higher than their true values simply to be done with the process. This may or may not be a problem in our approach. Further analysis and perhaps additional research will be needed to test this and possibly other concerns. In the meantime, we are taking a rather dim view of traditional bidding games.

The solution proposed by CBS is to let the respondents state their initial bids, perhaps with the aid of a payment card. Whether payment cards introduce a starting point bias of their own remains an issue for future research. The alternative is simply to let the respondent state an opening bid without the prompting of a payment card. This is like our Bidding Game 1. There, it will be recalled, respondents often did increase their bids both in the contingent and cash auctions. However, the increase in the mean bid was statistically significantly larger for the contingent auction. Stated differently, the process of iterative bidding in the contingent auction. caused people to bid money that they would not bid if the money was real. One study is obviously not definitive, but our evidence is contrary to the argument by CBS and others that bidding helps people research their preferences. We would think that it tends to encourage them to exaggerate their willingness-to-pay.

As a final note, the reader may wonder why all this is necessary, since final mean bids from Bidding Games 1 and 2 in the deer study were not significantly different than the results of the other mechanisms. Assuming that this result is replicated in later studies, it does raise additional questions about the efficacy of bidding games. Bidding rules out mail surveys and thus forces the use of more costly telephone and personal interviews. The ultimate conclusion may be that iterative bidding is not worth the trouble and expense.

4. Are Experimental Approaches The Key To Assessing And Improving CVM?

We agree that experimental approaches have great promise here. The experiment by Coursey, Schulze and Hovis (1983) (hereafter CSH), described in detail by CBS, is among the most interesting work done on CVM since its inception and illustrates well the potential usefulness of laboratory experiments. We hope that it will soon be one of many such studies. In this, we are in agreement with CBS.

Nevertheless, one has to wonder whether CBS are a bit one-sided in their emphasis on the virtues of laboratory experiments. Field experiments have a long established role in many disciplines, yet CBS

repeatedly imply that anything done outside the laboratory is second-rate science (see, for example, the discussion on pp. 85-86). In fact, our work does not warrant mention in their chapter on experimental economics, presumably because it was done entirely outside the laboratory. This is a very unfortunate precedent to set in this new area of economic research because it may divert attention from promising field experiments.

Perhaps research in aquatic biology will illustrate the need for combining laboratory and field experiments. University of Wisconsin limnologists have built a dike across the center of a lake in the northern part of the state. One side of the lake is to be acidified while the other will act as a control. Despite a long tradition of laboratory experiments and dozens of laboratory studies on the effects of acidity on aquatic organisms, many questions remain about what happens in natural ecosystems when pH is lowered from an external source. Such natural habitats can only be simulated to a limited extent and lab results are suspect because aquariums remain relatively artificial.

Does a similar problem exist for laboratory work on CVM? The virtue of the laboratory, as CBS emphasize, is a high degree of control. What they fail to bring out is that such control is gained by creating conditions that are highly simplified and highly artificial. A fish in the laboratory is still a fish, but the aquarium is not a wild habitat. Likewise, a human being in an economic laboratory experiment is an economic actor, but the laboratory situation is simplified and artificial. The result is that without field research there will always be questions about the applicability of results to the real world.

Consider again the CSH study. Again, our purpose is not to detract from their potentially very valuable contribution. Also, let us explicitly state that all we have for documentation is CBS' summary. We have not yet been able to acquire the papers that CBS cite. Nevertheless, CSH will help illustrate the limitations of laboratory experiments.

Suppose that a study of the economic losses due to air pollution in an eastern city is being planned. How much help would the HCS results provide? Could one generalize from bad tasting liquids to reduced visibility? The "commodity" in the CSH experiment was quite simple, while air pollution is complex, involving visibility, physical discomfort to eyes and the respiratory system, damage to public and private assets, and long-run health effects. Is behavior involving simple environmental "bads" in the laboratory necessarily indicative of behavior involving complex environmental bads in the real world? CBS do not describe the socioeconomic characteristics of the subjects in the CSH experiment, but presumably they would not be typical of a cross section of the population of the city in the air pollution study. Can we generalize from the laboratory subjects to the population in the applied study? The artificiality of the laboratory is also present in the way money enters in. Presumably -- although again CBS are unclear -- the CSH subjects were given some money to start with, at least those on the WTP side. Is it known what effects this had and whether people would behave differently in spending money out of their regular incomes?

Two points follow. First, in setting the agenda for future research, field experiments should go hand in hand with laboratory experiments. Second, all research results should be interpreted with care and laboratory results are no exception. Consider, for example, the use of HCS results to further discredit contingent WTA.

5. Should Willingness To Accept Be Abandoned In CVM Applications?

WTA has been a continual embarrassment to practitioners of CVM. Persistent, large differences between WTP and WTA have seemed at odds with theory and WTA values often seemed, in the eyes of the economic researchers at least, to be unrealistically large. Many studies have not even bothered to estimate WTA. Now CBS would use the results of the CSH experiment to drive a final nail in the coffin. Such a burial seems premature.

The deer-hunting study indicates that large differences between contingent WTA and contingent WTP are at least somewhat indicative of how people would behave if real money was involved. Further evidence is provided by CSH. There, the large differences persisted through at least four iterations of the actual cash auctions. Only after some unspecified -- at least by CBS -- number of trials did WTA collapse.

Objections to drawing general conclusions from this result come quickly to mind. Surely the arguments of the preceding section regarding differences between laboratory and real world conditions caution against automatically assuming that WTA will collapse under all conditions where CVM is applied. Furthermore, it should be noted that the CSH result was unexpected and somewhat mysterious. Assume that theory is correct in predicting that, for any individual, WTA and WTP will be equal, once equilibrium is achieved, except for the income effect. Assume also that CBS are correct in arguing that large observed differences between WTP and WTA during initial iterations of simulated markets and in CVM studies, reflect only the need of respondents to learn more about the market and their optimal strategy. Wouldn't learning be equally necessary for WTP? Wouldn't one expect a priori to see WTP and WTA converge in the middle, rather than convergence being solely the result of the collapse of WTA to roughly one-fourth of its mean value in early iterations?

Questions therefore arise about whether the CSH results reflect some basic economic principle with broad ramifications for all CVM studies or whether they only reflect something about the laboratory environment created by CSH. One can imagine, for example, high bidders seeing their low bidding competitors repeatedly drinking the SOA and making $10 or so. As the time in the lab comes to an end, such high bidders might reason that if they are going to make any money from the experiment they must underbid the competitor before the experiment ends. CBS do not provide enough information to even begin to judge whether such an "end-point bias" was operative. For example, did WTA taper off over several trials or collapse suddenly toward the end? How many trials on average were required? Were the lower values of WTA stable over several iterations after collapse or were they a transitory phenomenon? The CSH experiment is brimming with titillating possibilities for further research, but --- unless the papers are a great deal more persuasive than CBS -- it is hardly grounds for deciding to do away with WTA in all contingent valuation studies everywhere.

In fact, one might argue that recent research is grounds for more WTA research. From a theoretical standpoint, WTA is no more and no less to be preferred as a welfare measure than WTP. So long as one could appeal to Willig (1976) and Randall and Stall (1980) to say that the two measures were equal except for a probably small income effect, their joint existence was not a great concern. However, CSH and both of our experiments show that, at least during initial iterations, the differences are likely to be large, even though real money is involved. This phenomenon may have

important ramifications for welfare measurement.

Use our deer permits as an example. Assume for the sake of argument that the cash auction WTP and WTA values of $23 and $1,184 respectively, are "true" first-iteration values and that the problems of uncertainty alluded to in our discussion of the study do not exist. If the collapse of WTA, which CSH results lead us to expect, turns out to be a general principle, we would expect the WTA for deer permits to approach $23 eventually. However, for the 1983 hunt, the study subjects told us that, on average, it would take $1,184 to compensate them. If the 1983 hunt had been cancelled and it was somehow determined WTA was the appropriate welfare measure, surely the average loss would be $1,184 per permit, and not $23. Admittedly, if it were impossible to measure WTA, then that would make it impractical to use it as a welfare measure regardless of its theoretical niceties. However, in both of our experiments and in CSH, CVM worked about as well or perhaps better in estimating first -iteration WTA as in measuring WTP. Only in the long run is it necessary to worry about whether CVM is grossly overestimating WTA. In the short run, CVM estimates of WTA may well be relevant to policy and as capable of measurement as WTP.

6. Will The Application Of Vickrey Auctions Improve The Accuracy Of CVM?

Among the many themes developed by CBS, the advocacy of Vickrey or second-price auctions as a method to be employed in CVM studies stands Out as a dramatic departure from past thinking. Have CBS discovered a valuable new tool? We would rather think they have introduced a red herring.

The theoretical reason for needing a Vickrey format in actual Sealed-bid auctions is quite clear and convincing. The quote from Vickrey himself given by CBS on p. 89 makes the point well. Consider a situation where two men, A and B, are bidding for a single unit of a good. Assume that there is no collusion and that a first-price, sealed-bid auction is to be conducted. Suppose that CS(o) would be A's compensating surplus from consuming the good if he could get it for free. Let PA be his bid and PB be B's bid. Looking at the problem from A's point of view, he will not bid more than CS(o) since this would imply a welfare loss if he wins the auction. Thus, PA $\leq$ CS(o) must hold. Setting PA equal to CS(o) is not a particularly desirable strategy either. If A wins (i.e., PA = CS(o) > PB) then A will have to pay his full potential compensating surplus and be no better off. On the other hand, if PA is set sufficiently low that CS(o) > PA > PB then A can realize some net consumer surplus equal to CS(o) minus PA and be better off for entering the auction. On the other hand, if the outcome is CS(o) $\geq$ PB > PA then A will lose out and wish he had bid more. This is what A must balance in setting PA. He will tend to bid less than CS(o) to increase the gap between CS(o) and PA but he will also try to bid enough so that PA > PB. The exact bid will depend on his probability density function on PB. Still, the end result is a tendency to bid less than CS(o).

A Vickrey auction simplifies the problem greatly. The optimal Strategy will be to set PA = CS(o) If A wins (i.e., PA > PB), he pays only PB and realizes a net gain of CS(o) minus PB. If A loses, PB $\geq$ CS(o), so that A is no worse off. Thus, in a Vickrey auction, there is indeed an incentive to bid one's full WTP.

Note, however, that this is very different from what is done in a traditional CVM study. The hypothetical market in such studies does not ordinarily place subjects in a situation of bidding against each other for a

limited number of units of the amenity in question. Rather, the problem is simplified to one of determining at what price one would drop out of the market. In a way, this is more like an English auction where various participants drop out as they reach their respective maximum WTP's. In an English auction, as CBS point out, all people except the winner have an incentive to express bids up to their maxima. Thus, previous CVM studies have not failed to elicit maximum WTP simply by neglecting to have participants assume that they would actually pay the next lowest bid.

Of course, one could try to argue that it would be preferable in future CVM studies to have people assume they are bidding against others in a second-price auction for a limited supply of the environmental amenity in question. This would be theoretically as acceptable as the traditional approach, but not theoretically superior. Furthermore, the theoretical argument that people will reveal full WTP in a second-price auction depends critically on the assumption of expected utility maximization. It would hardly seem desirable to introduce uncertainty about what others will bid into CVM studies, given people's well-documented tendencies to behave in counter-theoretical ways under uncertainty. Also, since people are not familiar with second-price auctions, much more care would need to be exercised in designing survey instruments and even then there is risk of confusion. And, as CBS point out, several iterations may be required before respondents learn how to capitalize on the second-price format. It is not clear to us how one would structure the survey to provide a hypothetical situation conducive to learning what one would learn by actually winning and losing such auctions. Merely playing an iterative bidding game for a few minutes with an interviewer would not be much of a substitute for such experiences and could introduce additional problems. Repeated visits with reports of hypothetical auction results as proposed by CBS (pp. 98-101) sounds fine in theory but would be expensive, might cause respondent exhaustion, and would increase nonresponse problems as people became difficult to recontact. Without some way to encourage learning, the deer study indicates that a Vickrey format will not produce results significantly different from traditional results. Thus, the Vickrey framework would introduce additional uncertainty, respondent confusion, expense and complications into CVM applications with gains that are dubious.

Similar questions could be raised about the other departure from traditional CVM techniques suggested by CBS, the "tatonnement process" (pp. 100-101). Here, bidding and voting in successive iterations would occur until unanimity about payment and pollution allocation is achieved. Such tatonnement processes would allegedly "out perform" (p. 101) more traditional procedures. To sustain this argument, however, at least two assumptions must hold. In traditional CVM applications to commodities with true public good characteristics (e.g., visibility), normal procedures in essence ask respondents to pretend that the commodity is a private good. Thus, for the procedures advocated by CBS to be necessary, it must first be assumed that study subjects are unable or unwilling to imagine the commodity as a private good. This assumption seems doubtful given the lack of evidence of free riding described by CBS, Even if the first assumption does hold, one would also have to assume that going through a hypothetical Grove-Ledyard procedure would cause respondents to reveal their true preferences and values. If, contrary to present evidence, they are already free riding, why should they change in a hypothetical situation? The-alternative of increased cost, increased confusion, and lower response rates

seems a more likely result of attempting such procedures.

7. Is Attitude-Behavior Research Relevant to CVM Research?

Beginning with our first publication on the goose study (Bishop and Heberlein, 1979) , we have attempted to introduce economists to the research by social psychologists on attitudes and behavior. We argued that CVM expressions of WTP and WTA are, in psychological terms, "attitudes", while actually buying and selling things is "behavior". In questioning whether contingent values are accurate, economists are, in a sense, asking whether attitudes (expressions of WTP) correspond to behavior (how people would behave if a real market was created). A major result from the attitude-behavior literature was introduced. In general, the relationship between measured attitudes and actual behavior varies greatly and in many cases is quite low. In the current context, this serves as a warning against assuming automatically that people actually will pay or accept what they say in a survey they will pay or accept.

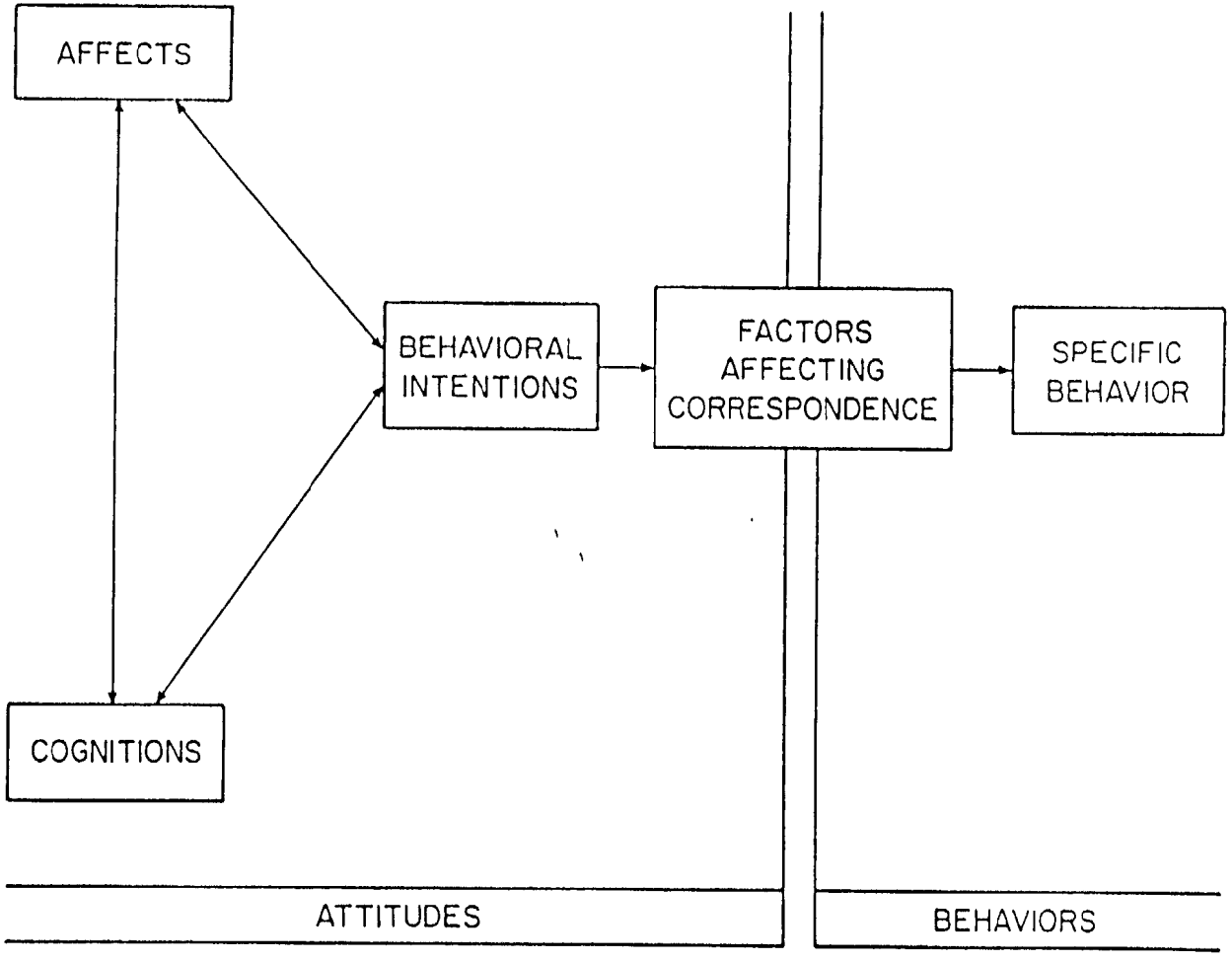
We certainly underestimated the barriers to interdisciplinary communication. Our proposal that economists consider the attitudes-behavior literature has met with indifference or hostility. CBS are no exception. Nevertheless, we continue to believe that this material is relevant and that economists are the losers for ignoring it. Allow us to attempt to make our case clearer.

An attitude is a mental state relating to some object. That is, a person has an attitude about something. The object may be very general as in the case of environmental attitudes or very specific as in one's attitudes about one's spouse (Heberlein, 1981). Attitudes generally have three related components. The "cognitive" component refers to dispassionate facts and beliefs. For example, a person might say that the water in Lake X is clean. Second is the "affective" component. Affects have to do with the evaluative and emotional aspects of attitudes. A person might say "I like swimming in clean lakes." The third component is "behavioral intentions." Continuing the example, a statement of behavioral intent might be, "I plan to swim in Lake X this summer". For the most part, responses to contingent valuation questions are, to the social psychologist, statements of behavioral intention. In a WTP question, people are saying that if a market existed for the amenity in question, their intention would be to pay certain stated amounts. No actual behavior has taken place, but people have expressed an intention to behave in a certain way.

As in any discipline, social psychologists adapt the terminology to their own needs. In the present case, Ajzen and Fishbein (1977), as cited by CBS, use the term attitude more narrowly to refer to the affective component only and apply the term behavioral intention separately. Terms do not really matter here, so long as confusion is avoided. The ideas are the same. Our terminology is more consistent with the bulk of the literature and we will continue to use the term attitude in the broader, more all-encompassing sense.

The left-hand side of Figure 9.1 illustrates the linkages between the three components of attitudes. In everyday language, when we "think about" something, the three components interact. For example, liking clean lakes (an affective component) may, over time, encourage us to gather information about which lakes are clean, building the cognitive component. The arrows run both ways. For example, learning that Lake X is suffering from

Figure 9.1: Schematic of Attitude-Behavior Relationships



declining pH due to acid rain (a cognitive component), I might decide that I only like clean lakes that are also unaffected by acid rain. My behavioral intentions toward Lake X may change as a result.

Social psychologists draw a very basic distinction between attitudes and behavior. This is depicted by the vertical double line in Figure 9.1. To observe that Lake X is clean or to state that one likes swimming is not the same as actually going swimming. Even stating plans to go swimming is not the same as actually doing so. Only when one actually gets in the water is the link between attitudes toward Lake X and behavior with respect to Lake X completed. Behavior is something that can be observed in the real world. Attitudes are not directly observable, but must be inferred, usually from survey responses.

These relationships were clear in the goose study. Recall how commitment came into the equations for both the hypothetical and cash offers. Commitment expressed how the subjects felt about goose hunting. An element of behavior intention may also have been present in committment. The cognitive component included new knowledge in the form of a real or hypothetical offer from the University. Both commitment and the amount of the offer interacted to influence the economic behavioral intention (yes, I would sell or no, I would not). However, the cognitive component was different in the two treatments. In one case, the respondents knew the offer was real while in the other they knew it was hypothetical. Thus, there was a divergence between the behavioral intentions expressed in the contingent market and behavior in the simulated market. Most probably they didn't purposely mislead us, but the different cognitive components lead to a different set of interactions as they thought about the offers. Commitment tended to have more influence for hypothetical offers; dollar amounts had more influence for cash offers; and the result was a substantial difference between behavioral intentions and behavior.

That attitudes do not always predict behavior should not be surprising. Focusing attention on the box near the center of Figure 9.1, many factors affect attitude-behavior correspondence besides attitudes. An interesting example can be drawn from the CSH experiment. Consider those who at the outset said that they would require almost $10 on average to taste SOA, based on a verbal description. However, in Part II they tasted the stuff without being paid anything (at the margin) to do so. Here is a simple case where attitude ("I'd have to be paid $10 to taste the stuff-") and behavior (tasting free) did not correspond. Obviously, there was an additional factor at work. They had signed up for and were presumably being paid to participate in an experiment and behaved as they did because tasting SOA was part of the experiment.

As noted already, one of our goals in introducing all this was to warn economists that attitudes, including behavioral intentions, are not necessarily the same as behavior. A second reason for linking economics and social psychology is becoming increasingly clear. The attitudes-behavior work is a rich source of both theoretical and empirical insights of direct relevance to CVM studies. Let us attempt to further support this assertion.

We will illustrate application of attitude-behavior concepts by referring to our own current research on acid rain. (Bishop and Heberlein, 1984) Reductions of 50 percent in sulphur emissions from power plants east of the Mississippi may cost as much as $5 billion per year. This raises questions about the magnitude of associated benefits. In the aquatics

area, economists can estimate the value of fishing losses in areas such as the Adirondack Mountains. In fact, such studies are in progress. Both economists and noneconomists are asking whether such use values alone will fully capture the economic losses associated with acid rain. Thus, our own work is focusing on the "non-use" or "intrinsic" values.

The terminology of intrinsic values has not been agreed upon by all resource economists. In our thinking, intrinsic values fall into two broad categories, option value and existence values. Option value is too complex to be dealt with in any detail here. It must suffice to say that option value is a premium, positive or negative, associated with uncertainty about future use of the resource (Bishop, 1982; Graham, 1981; Smith, 1983; Freeman, 1984). Existence values, on the other hand, have to do with values that people would still hold even if use were constrained to zero. The concept can be traced back to Krutilla's (1967) landmark article on conservation economics. Other conceptual work appears in Krutilla and Fisher (1975), Mitchell and Carson (1981), Randall and Stoll (1983), and Desvousges, Smith and McGivney (1983).

Elsewhere, (Bishop and Heberlein, 1984) we have argued that existence benefits for reductions in acid deposition rates could be positive for several reasons, including: (1) bequest motives; (2) benevolence toward relatives and friends; (3) sympathy for people and animals affected by environmental damages; and (4) feelings of responsibility for environmental damages caused, for example, by use of electricity generated by coal-fired power plants.

Existence benefits from reduced sulphur emissions could, of course, be estimated using CVM. Even if small on a per household basis, such benefits, when added up over millions of households, could be quite large. In fact, we suspect that, given the widespread concern about acid rain and the relatively limited extent of documented current and probable, near-term future damages, existence benefits estimated using CVM will dwarf use benefits. The direction that the economic scales tip could well depend, at least over the next decade or two, on whether the existence benefits have credibility. Thus, economic conclusions about a major national policy issue may depend on whether CVM estimates of existence values are accepted as valid or not.

Empirical assessment of the validity of contingent existence values will not be easy. Field experiments like those involving hunting permits do not appear promising. Laboratory research might be feasible, but experimental designs are not obvious to us, This is where the attitude-behavior research could prove useful.

The question is a relatively straight-forward one of attitude-behavior relationship. Would people expressing the behavioral intention of paying a certain amount for reduced acid emissions actually pay that amount if a market for existence were created? There is a large body of research on the conditions favorable to attitude-behavior correspondence.

As CBS recognize, the strength of the attitude-behavior relationship can be assessed by looking at the specificity of the behavioral intention measure. Drawing on Fishbein and Ajzen (1975), behavioral intentions are stronger predictors of behavior the more specific they are about targets, actions, context, and timing. Target specificity has to do with how definite survey and interview questions are about the actual target of behavior. For example, one would expect a question about existence value of fish in a certain Adirondack region to be more highly correlated with

behavior in a real existence market than very general questions about vaguely defined acid-rain damages. In the present context, action and content specificity have to do with stating whether payments will be higher monthly utility bills, taxes, prices or other modes of payment, rather than asking vague questions about "WTP". Timing is important because attitudes change. The shorter the time between attitude statement and actual behavior the better is the relationship between the two. Thus, one would expect contingent valuation questions to predict better, other things being equal, the more specific they are about timing of hypothetical payments and the shorter are the time horizons designed into the contingent valuation mechanisms. Ex ante, the researcher can and should take these factors into account in designing studies. indeed, the better practitioners of contingent valuation are already doing so. The point here is that after the contingent valuation mechanism has been designed and applied, one can be more confident about validity, the more successful one was in designing specificity into the mechanism.

Going beyond contingent valuation mechanism design, other data can be gathered during the survey process to evaluate the possible extent of attitude-behavior relationship, By definition, expressions of WTP for the existence of reduced acid deposition rates involves altruistic behavioral intentions toward the environment. Previous research, dealing with environmental altruism with respect to littering, early use of lead-free gasoline, and energy conservation (Heberlein, 1975) has isolated two factors that are particularly important in activating actual behavior consistent with such altruistic behavioral intentions. These are awareness of consequences (AC) and acceptance of personal responsibility (AR). AC has to do with cognitive understanding of ecological effects and particularly awareness of effects on other people, AR refers to how strongly people believe that they are personally to blame for environmental degradation. People with low AR may place blame on other people, corporations or the government. People with high AC and AR have a stronger tendency to carry altruistic behavioral intentions toward the environment into actual altruistic acts, while people with low values for either or both tend to have low correlation between attitude and behavior.

The concepts of AC and AR match well with economic intuition that valid existence values must be related to bequest, benevolence, sympathy, and/or other motives discussed previously. For example, a person who expresses a high existence value for acid rain reductions based on bequest motives is implying (1) awareness that acid rain damages will affect future generations and (2) that he or she is personally responsible for reducing these effects.

Thus, a clear direction for acid rain research emerges. It is important not only to measure people's contingent existence values, but also the major reasons why they may be expressing those values. Cognitive attitudes about acid deposition and its consequence should be measured. Attitudes toward future generations and the stewardship role of the present generation should be examined. Knowledge about and sympathies toward relatives and friends who might be affected by acid rain may also be important. Questions relating to actual altruistic behavior toward the environment and other "causes" (e.g., recycling cans and bottles, membership and level of activeness in environmental organizations and charitable contributions expressing sympathy for people and animals) should be included in the survey instrument, If bid equations show significant positive

relationships between CV existence values and these variables, this would support the hypothesis that the prerequisites for carrying altruistic behavioral intentions into action are present. If no relationship exists, doubts would arise about the prospects for strong attitude-behavior relationship. The validity and, hence, policy relevance of the economic values would then be more questionable.

CBS are very pessimistic about general prospects for accurately measuring existence values using CVM. Our own remarks should not be interpreted as indicating that we are taking lightly the concerns they express. It is particularly disturbing that there is so much scientific uncertainty about the nature and extent of acid rain damages. The presence of this uncertainty must surely be incorporated into the valuation process. Preference reversal and other observations from experiments involving uncertainty are cause for concern. Still, if conditions for high attitude-behavior correspondence are fulfilled, some grounds would exist for arguing that legitimate economic values are being established at least to a rough approximation.

Hopefully, the acid rain example illustrates that the attitude-behavior literature is of value to CVM researchers. In fact, CBS can find substantial empirical support for many of their conclusions in that literature. For example, their first two ROC's (familiarity with the product and prior experience) appear to be quite consistent with social psychological research results.

C. OUR ASSESSMENT OF CVM.

Tony Scott (1965, p. 37) once remarked, "Ask a hypothetical question and you get a hypothetical answer." We came to CVM research with the same cynicism. To some degree, our research has added empirical support to Scott's assertion. Hypothetical bias does appear to be an inherent weakness of CVM.

Still, we have been surprised a how well CVM does work. In the goose study, the dollar amount in take-it-or-leave-it WTP and WTA offers was consistently the most powerful variable in predicting response, always coming into the logit equations with the expected sign and with significance at the .01 level. Most of our respondents certainly understood what was being asked of them and there was a tendency to respond in the same way they would in a real market, albeit in an imperfect way. Similar conclusions seem to follow from the deer study. On the WTA side, the hypothesis that hypothetical and cash offer means were equal could not be rejected at the .10 level. The CVM mean for WTP was significantly higher, but was certainly not outrageous. Deer management decisions in Wisconsin would probably not be greatly different if based on the CVM estimate of $40 per permit rather than the cash auction mean of $23.

Thus, while CVM appears to be biased even under the best of circumstances, the degree of bias does not appear to be sufficiently high to rule out use of the results in public decision-making. While asking a hypothetical question does elicit a somewhat hypothetical answer, it is also true that if a well-constructed question is asked, people try to give honest answers. This, in our judgement, makes CVM promising.

To fully capitalize on this potential will require a new commitment to methodological research. Past research in this area has not been as conducive to real methodological progress as it might have been for two reasons. First, it was probably necessary for CVM to go through a prescientific stage. Most of the history of CVM brings to mind children with a chemistry set pouring chemicals at random into a test tube to see what will happen. (Perhaps the most recent installment is to "stir in" a Vickrey auction.) Second, there has been very little truly basic research on CVM. Most of the research has had to justify its existence by claiming to address real-world problems. Methodological research had to be done as an add-on to these applied studies. It is little wonder that after 20 years, we are still debating such basic issues as whether iterative bidding improves accuracy.

CVM has shown itself sufficiently promising to warrant a major basic research effort. CBS are quite correct in suggesting that experimental techniques are the key, particularly if they will admit the importance of field as well as laboratory studies. Their hard-headed insistence On stating testable hypotheses may help us get beyond the "chemistry set" approach. The ultimate goal ought to be to go beyond error bounds and counting significant digits to actually overcoming hypothetical bias through calibration.

Agencies such as EPA that have a large interest in developing CVM techniques need to recognize that such basic reseach probably will not be feasible in the context of the policy issues of the day. To address such policy issues more effectively, funds need to be set aside for studies in settings that are more ideal for methodological research. Such research may have to deal with commodities such as SOA and hunting permits before we can do a better job on acid rain in the Northeast and air pollution in Los

Angeles.

To pause and examine the state of the art after 20 years and millions of dollars worth of research is worthwhile. A great deal has been learned about CVM, but so much is unknown even now. We do know that CVM is the most promising technique for applying an economic yardstick to many of the nation's seemingly most valuable environmental and resource commodities. Enough positive evidence has accumulated to warrant a major investment in full development of the contingent valuation method.

ENDNOTES

Chapter IX

1) Contribution to "Valuing Environmental Goods: A State Of The Arts Assessment of the Contingent Valuation Method" by R.G. Cummings, D.S. Brookshire, and W.D. Schulze. Research was supported by the College of Agricultural and Life Sciences, University of Wisconsin - Madison; Resources For The Future, Inc.; the Graduate School of the University of Wisconsin - Madison; and the Electric Power Research Institute. The Sandhill deer permit valuation study was done with the help of the Bureau of Research and the staff of the Sandhill Wildlife Demonstration Area, Wisconsin Department of Natural Resources. Glen Anderson, Kevin Boyle, and Michael Welsh made many helpful comments on an earlier draft. All errors are the sole responsibility of the authors.

X. ON ASSESSING THE STATE OF THE ARTS OF THE CONTINGENT VALUATION METHOD OF VALUING ENVIRONMENTAL CHANGES.

By
A. Myrick Freeman III
Department of Economics
Bowdoin College

A. INTRODUCTION

The subtitle of the report we are discussing is "A State Of The Arts Assessment ...". This is a felicitous choice of words, I think, because the impression I get from reading the Assessment is that the design and implementation of a CVM survey is still more of an art than a science. Although we have learned a lot about the problems involved, we must still rely to a large extent on the good judgement of the researcher in dealing with such problems as incentives to strategic behavior, starting point bias, the best way to describe the commodity being valued, the choice of a payment vehicle, and so forth. Also, as is the case with the arts, the criteria for evaluating CVM research are not well defined. Judgements concerning the usefulness of the technique and the validity of individual CVM surveys appear to be to a large extent subjective. Different people reach quite different conclusions about the merits of the technique as a whole and individual studies.

This Assessment is valuable, at least in part, in that it attempts to move beyond subjective and impressionistic judgements and to place the evaluation of the CVM technique on an objective, scientific foundation. It does this by focusing attention on the question of the accuracy of CVM measures of value, by formulating hypotheses about factors that might influence the accuracy of CVM responses and by reviewing the evidence about these hypotheses that can be gleaned form the accumulated body of CVM data.

In what follows, I, too, will focus on the question of accuracy. I will first discuss the forms for evaluating the CVM, one of which is accuracy. I will then discuss the two forms of inaccuracy in CVM measure, bias and random error. I will then discuss the author's concluding assessment and provide more conclusions of my own. I will also provide some specific comments on points where I take issue with the authors' analysis. My assignment was to provide two assessments: one of the authors' report and one of the CVM itself. I have chosen not to organize my response along these lines. Rather in what follows, my ideas concerning the CVM are intertwined with my comments on the authors' assessment.

Before turning to a detailed discussion of the Assessment, I want to point out what I think is a serious limitation in the scope of the Assessment. The authors hereafter referred to as CBS) restrict their discussion to those contingent choice methods designed to elicit directly a monetary valuation of the environmental good. There are at least four types of what I would call contingent choice mechanisms which have in common the objective of eliciting information which can be used to determine a monetary value by posing to individuals hypothetical or contingent questions of the form "What would you do if ...?" or "How much would you pay if ...?". The first type, which is analyzed in this Assessment, asks for information on the monetary value that the individual attaches to a specified change in the quantity or availability of the environmental good.

The second type, which is more relevant for the analysis of private goods demand, asks the individual to indicate the quantity she would wish to purchase at a specified price.

The third type is the contingent ranking method. With this technique individuals are given a set of cards, each card depicting a different set of conditions with respect to the use of the environmental good, including differences in the level of provision of the environmental good itself and perhaps different prices or admission fees for use of the resource. Individuals are asked to place their cards in order of preference. Marginal rates of substitution and monetary values can be inferred from these rankings. Examples of contingent ranking studies include Desvousges, Smith, and McGivney (1983) and Rae (1981).

Finally, individuals might be asked how they would alter activity patterns such as rates of visitation to different recreation sites in response to changes in the level of provision of an environmental good at one site. In some circumstances it may be possible to infer monetary values from reported changes in activity levels. Examples of this technique include the willingness to drive model of Knetsch and Davis (1966) and the site substitution model of Thayer (1981). 1/

A comprehensive assessment of contingent choice methods would include a consideration of whether any of these techniques has any advantages over the CVM in terms of ease of implementation, reduction in the various forms of possible bias, or accuracy of the inferred valuations. For example, appropriate strategies may be more difficult to discern in the case of contingent ranking or site substitution models, thus reducing the likelihood of strategic bias. And both of the latter models appear to avoid starting point problems. But since they ask "What if ..." questions, they may be subject to what has conventionally been termed hypothetical bias. At least it seems to me that these are important questions to take up in a comprehensive and full assessment of contingent valuation methods.

B. CRITERIA.

Any assessment of a technique for eliciting a valuation must be carried out in terms of one or more agreed upon criteria or standards. CBS are aware of this as their discussions of the need for standards for providing or disproving hypotheses (pp. 9-10) indicates. But I would have thought that CBS would devote more attention to the criteria to be employed in this assessment at the beginning of the paper. It is not until pages 147-150 that we find an explicit statement of the criterion they propose to employ in the following assessment. There they say:

> "Thus, the general criterion against which to assess the CVM becomes becomes clear: the extent to which the CVM institution, and preference revelations drawn therein, is comparable with the market institution and preference revelations drawn therein." (pp. 148-49)

Unfortunately, I find this statement somewhat confusing. It is not clear which is thought to be more important, the comparison of the institutions or the revealed preferences and valuations. And the statement does not distinguish between individual and aggregate responses. I want to offer an alternative statement of what I think the principal criterion for an assessment should be. I agree that the principal criterion should be the accuracy of the resulting measure of value. By accuracy I mean the degree of correspondence between an individual stated value (or his revealed value in the forms of contingent choice methods) and his true value. It is important to make it explicit that individuals' responses are at issue so as to distinguish between problems in eliciting accurate values on the one hand and sampling from a population and aggregating across individuals on the other. Sampling and aggregation problems are not at the heart of the controversy over CVM Finally, the nature of the CVM instrument should not be part of the criterion. The CVM institution itself is of direct significance only to the extent that it facilitates the revelation of true or accurate values.

I want to spend a little more time to consider just what I mean by the "true value". According to the standard definition of a compensating measure of value, the true value is that sum of money which the individual would give up (or accept) to restore himself to his original utility level given an increase (or decrease) in the quantity of the environmental good. 2/ In other words, the true value is the income/environmental good trade-off which maintains the individual on his original indifference curve. It is conventional to assume that individuals have well defined preference orderings and that they know the shape of their indifference curves. Thus, if we observe an individual to accept a trade-off between income and some other good, we believe that he has revealed something about his preference ordering and the shape of his indifference curves. But the inference that revealed trade-offs reflect true valuings or preferences is correct only if individuals do in fact have full knowledge of their preference orderings.

Suppose that due to a change in relative prices or income or the introduction of a new good, an individual has an opportunity to choose from among a set of consumption bundles that are unfamiliar to her, that is, which she has had no prior experience with. It seems plausible that she might experiment with several different consumption bundles before settling

into a new equilibrium position. This experimentation can be viewed as an effort to explore an unfamiliar part of her preference ordering. We can only accept revealed preferences as reflecting true preferences after this exploration has been completed. Therefore I want to define the true value of the environmental good as that substitution between income and the environmental good which we would observe after repeated trials or opportunities for the individual to alter her consumption position.

The reason that we have confidence that individuals reveal true preferences in their market behavior is that they have many opportunities to modify their choices in light of what they learn about their preferences and the characteristics of goods. Similarly many economists, myself, included, have expressed confidence in measures of the value of environmental goods derived from hedonic price models and travel cost models because they reflect choices made by individuals who have an opportunity to learn from their experiences and modify their choices accordingly.

A measure of an individual's value of a change in the provision of an environmental good is accurate to the extent that the measured value corresponds to the true value as defined above. Inaccuracies or errors in the measured values produced by a given technique or instrument can have two components. The first is a random component or random error reflecting some structural problem or fault in the technique. In the next section, I consider sources of bias or systematic error in CVM measures. In section IV I discuss possible random errors in the CV< technique. But before turning to discussion, I want to mention two additional criteria that may be relevant in the choice of a technique for estimating values for environmental policy making.

One criterion is how much information does the technique provide on the individual's preferences or valuation for the environmental good. Ideally, we would like to recover the individuals inverse demand function for the environmental good so that measures of value for the individual can be calculated for a wide range of changes in the quantity of the environmental good under a wide variety of conditions. An individual's response to a single willingness-to-pay question is an estimate of the integral of the compensated inverse demand function over the range given by the postulated change in the environmental good. But this does not provide enough information to make reliable estimates of the individual's value for larger or smaller changes in the environmental good. The single response can also be interpreted as one point on a Bradford bid curve Bradford, 1970). The responses to additional questions postulating different changes in the environmental good yield additional points on the Bradford bid curve. If sufficient information can be obtained to estimate the bid curve, then the income compensated inverse demand curve can be recovered by differentiation.

The other criterion is cost. Some people have suggested that contingent valuation experiments are easy to set up and provide an inexpensive source of valuation data (e.g., Randall Hoehn, and Tolley; forthcoming). From my own observation of the design and implementation of the Vanderbilt survey of the benefits of hazardous waste regulations I am not convinced that CVM surveys are either easy or cheap. It seems likely that the cost of a survey is an increasing function of its accuracy. Accuracy is likely to be a function of both sample size and the effort devoted to reducing sources of bias and measurement error in the design of the survey instrument. We need to know more about the cost and accuracy of

CVM instruments as well as the cost and accuracy of alternative measurement techniques where they are available before we can advise analysts concerning the selection of measurement techniques in particular circumstances.

C. BIAS.

In this section I will discuss strategic biases, starting point bias, information bias, and vehicle bias. Since I lean toward the view that the hypothetical nature of the CVM instrument is more likely to lead to random measurement error than to bias, I defer my discussion of hypothetical bias to Section IV.

1. Strategic Bias.

The first source of bias to consider is that resulting from conscious attempts by individuals to influence either their payment obligation or the level of the provision of the environmental good through their stated valuations. One form of strategic bias arises from the efforts of respondents to "free ride", that is, to reduce their repayment obligation by stating low values. Others involve efforts to influence the level of provision of the environmental good by stating artificially high or artificially low values. It is important to note that the opportunity for strategic behavior arises only when the valuation questions are asked in a setting in which it at least appears that the actual outcome will be affected by individuals' responses, that is, in other than the purely hypothetical or contingent market setting.

My own view is that strategic behavior should not be a significant problem in carefully designed CVM instruments. This judgement is base on three considerations. The first is the absence of strong evidence for free rider behavior in experiments designed to test the free rider hypothesis (Smith, 1979; Marwell and Ames, 1981). The second is the fact that most CVM instruments do not offer obvious opportunities or incentives for attempting to manipulate the outcome. And finally, visual inspection of the distributions of bids does not suggest strongly biased response, although this is admittedly a weak test.

Designing CVM instruments to avoid serious strategic bias may involve an element of art or at least judgement on the part of the analyst. The art involves providing a realistic description of the environmental good to be valued and policy scenario while making it clear that real world policy decisions are unlikely to be directly affected by the values revealed by the survey. There may be some situations which invite inflated responses from some groups, in which case CVM surveys would not be likely to provide reliable data. For example, suppose there was a widely publicized proposal to dam and flood the Grand Canyon. A CVM survey of visitors to the Canyon asking their willingness to pay to preserve the Canyon would offer people an opportunity to register their disapproval of the proposal. Thus CVM surveys may be less reliable when they deal with highly politically charged policy questions.

There is a problem which is somewhat related to strategic behavior about which I cannot be sanguine. That is the significant number of refusers and protest zero bidders that are often found in CVM studies. The person who refuses to state a monetary value on the grounds that it is unethical to do so or that he has an inherent right to the environmental good must be dropped from the sample when mean bids are calculated. If a person bids zero on the grounds that he has an inherent right to the good, the bid is not an indicator of his true valuation. Thus an effort should be made to distinguish protest zeros from true zeros so that the former can be dropped from the sample, too. It seems plausible that at least some refusers and protest zero bidders are using a noneconomic means of expressing high economic values. If this is so, then there is a kind of

self selection bias at work resulting in a downward bias in estimated sample mean bids. On the other hand Carson and Mitchell (1984) suggest that many nonrespondents are poor and have low levels of education. If their true values are relatively low, then the mean of remaining responses is an upwardly biased measure of the true mean value.

This analysis suggests three considerations in the design of CVM instruments and the reporting of results. First, efforts should be made to word the CVM question so as to minimize the numbers of protest zeros and refusers. Pretesting of survey instruments should help to detect those aspects of questions which stimulate protest behavior. Second, all zero bidders should be queried so as to identify protest zeros; and protest zeros should be cropped form the valuation sample. 3/ And third, the proportion of the original sample which is dropped because of refusal or protest zero bids and the characteristics of other nonrespondents should be reported as an indicator of the possible bias in responses due to self selection.

2. Starting Point Bias.

There is ample evidence that starting point bias can be present when a staring bid is announce by the interviewer and the offer price is adjusted upward or downward until the respondent agrees on the stated value. Also there is evidence that when the respondent is asked to name a value for willingness-to-pay, he can be induced to adjust this upward by a series of iterative questions. There are a couple of ways in which the starting point problems might be dealt with effectively.

First, consider the starting point bias problem in its simple form. If the mental mechanisms which lead to starting point bias are such that the bias is a function of the absolute value of the distance between the starting point and the individual's true value, and if the upper and lower starting points are equidistant from the true sample mean value, then the two biases can be made to cancel out. With the sample equally divided between low and high starting points, the best estimate of the true value is the mean of all responses.

Another possible approach is to derive an iteration procedure from the "bracket and halving" procedure used to adjust naval gunfire to strike a target. The procedure would be to announce a starting point chosen at random for each respondent within the range of likely values. This offer would be adjusted in the appropriate direction by a large enough step so as to bracket the individual's likely true value. Successive adjustments would involve halving the interval between the two preceding offers as appropriate until the individual agreed on the stated value. This procedure is designed to close as rapidly as possible on the true value, thus reducing the likelihood of a boredom effect. Also choosing the initial bid at random avoids the indicative effects of nonrandom starting points. Thus, even if each individual's response has a systematic error related to his starting point, these errors can be made random across individuals so that aggregate value are unbiased.

3. Information Bias.

Two kinds of information bias have been discussed in the literature. One refers to the effect of providing information on values and costs (for example, the cost of providing the environmental good, the costs and/or values of other kinds of public goods, or bids offered by other respondents). If this kind of information is provided, it would appear to lead to bias through a kind of indicative effect akin to that leading to starting point bias. For this reason it seems that this kind of information

should not be provided to respondents.

The second type of information bias is said to result from changes in the information provided to individuals about the environmental good itself. Evidence that individuals' bids can be changed in systematic ways by changes in the description of the environmental good should be taken as favorable to the CVM. This evidence indicates that people use the information provided to form a perception of the environmental good and base their valuation responses on their perception. I think that two conclusions can be drawn about the design of a CVM instrument and the interpretation of its results. First, it is important to provide a clear and meaningful description of the environmental good of concern. Here, too, the art of CVM instrument design is important. Second, statements about the results of CVM measures should take the following form: "The value of the environmental good as described in the CVM instrument is $X." This qualifying phrase makes is clear that what is being valued in the CVM exercise is the environmental good described to the individual. The relevance of the CVM results for valuing the outcome of a real world environmental policy depends upon the degree of correspondence between the environmental good described to individuals and the proposed real world environmental change.

4. Vehicle Bias.

Vehicle bias refers to systematic differences in responses depending upon the postulated means of collecting payments from individuals. Some studies find systematic differences between payment vehicles while others do not. Interpretation of those studies which do find differences is hampered by our inability to state which payment vehicle, if any, provides "true" values and which payment vehicles lead to bias. Here again, the artful instrument designer may be able to learn from an examination of earlier studies how to specify payment vehicles so as to minimize vehicle bias. One approach to learning about vehicle bias might be to ask attitude questions about various payment vehicles in an effort to identify those which do not invoke negative attitudes in given circumstances.

5. Summary.

I have argued that the problem of strategic bias and starting point bias can probably be minimized by the careful design of the survey instrument. Information bias that results from a divergence between the true environmental good and the description provided to respondents probably should not be termed a "bias". It is the description that is biased, not the valuation of what is described. Vehicle bias and self-selection bias resulting from protest zeros and refusers are more troublesome. The likely presence of vehicle bias can be identified if two different vehicles are tested in the same instrument. But we lack an objective means of determining which, if any, of the vehicles indicates the true value and therefore the direction of bias is unknown. It seems likely that self selection will bias willingness to pay values downward and that this bias will be stronger the larger the proportion of refuses and protest zeros in the original sample. But this is a conjecture. In the absence of an independent way of estimating individuals' true values, this conjecture cannot be tested.

D. RANDOM ERROR

In this section I will discuss what I in my book (Freeman, 1979, pp. 97-99) called the problem of accuracy as distinct from bias (I now regret that choice of terminology) and what others have called hypothetical bias (which may or may not be bias, but results from the hypothetical nature of the CV instrument). In my earlier treatment of the problem I argued that the accuracy of a revealed value (that is, the degree of correspondence between the revealed value and the true value) depended on the time and mental energy devoted to the decision process. Since time and mental energy are costly, increasing accuracy comes only at increasing cost to the individual. The benefit of accuracy is the avoidance of foregone utility due to nonoptimal choices. I argued that in the hypothetical settings of the CVM. since individuals did not have to live with the consequences of their choices, the incentives to make accurate responses were weak. Although I was not explicit on this point, I believed that those errors would be random with zero mean.

I now believe that there is another element to the individual choice problem in a hypothetical setting which can lead to potentially large random errors in individuals' reported values over and above those associated with the absence of incentives or time. This element has to do with individuals' familiarity with the environmental good and their experience with changes in its level of provision. Note that these two terms, "familiarity" and "experience," are used by CBS in defining the reference operating conditions for the CVM)p. 199). Their treatment of this set of questions in Chapter VI has helped clarify my thinking on this issue.

It is conventional to assume that individuals have well defined preference orderings over all goods, including public good and environmental goods. We assume that these preference orderings can be represented by utility functions of the $U = U(X,Q)$ where X is a vector of private goods and Q is the level of an environmental good. It is conventional to assume that individuals have accurate knowledge of their preference orderings over the full range of bundles in their choice sets. My key assumption is that individuals have better or more accurate knowledge of their preference orderings in the neighborhood of those consumption bundles they have actually experienced. If shifts of the budget constraint induce an individual to move into an unfamiliar region of his preference ordering, he is likely to make mistakes in his initial choices of consumption bundles, that is, initial choices will not be accurate reflections of the true underlying preferences. Only after the individual has had a chance to learn about or gain experience with this region of his preference ordering and correct any initial mistakes in choices can we infer true values from revealed choices. This is what I had in mind above in defining a true value as one reflected in repeated choices and implying the absence of regret.

Now suppose that the level of the environmental good has been at Q^* throughout an individual's life. It is reasonable to believe that the individual knows his marginal rates of substitution between Q and other goods in the region of Q^*. But for levels of Q substantially different from Q^*, the individual may have only a vague idea of his marginal rates of substitution between Q and other goods. This means that given a substantial cange in Q, the individual's initial adjustments in the quantities of private goods purchased may be different form the consumption

bundles finally settled upon after gaining experience with the new level of Q. I hypothesize that the difference between the initially revealed preferences and the final or true reveled preferences will be random and will be on average larger, the larger is the change in Q.

A corollary of this hypothesis is that CVM responses to questions about small changes in Q in the neighborhood of Q* will be more accurate than CVM responses to questions about large changes in Q, especially if the individual has had no prior experience with the alternative postulated level of Q. Also, it seems to me, these errors should be random with zero mean, but more on this point below.

CBS must have had a model of choice and learning of this sort in mind when they produced their reference operating conditions 1 and 2 (p. 199). In this sense my analysis is consistent with theirs. However they did not explicitly adopt this framework in their Chapter V "Imputing Actual Behavior form Choices Made Under Hypothetical Circumstances." In this chapter CBS make a valuable contribution in that they attempt to deduce testable hypotheses about the relationships between hypothetical values and true values from various arguments that have appeared in the literature, and to subect these hypotheses to empirical tests based on existing CVM data However I think CBS are not entirely successful in this effort. But this is at least in part because the arguments that they are evaluating have not been well formulated, and in part because of the difficulty in finding operational measures of some of the theoretical constructs.

For example, CBS quote me on the implications of the absence of incentives to accuracy, and then formulate the null hypothesis: values revealed when incentives to accuracy are present will be equal to values revealed with no incentives to accuracy. In my formulation, the incentive to accuracy was the avoidance of the foregone utility associated with nonoptimal choice. But they equate incentive with a requirement to make the offered payment, so that the null hypothesis becomes: values revealed with no requirement for payment will be equal to value revealed when payment is required. This is clearly a different hypothesis. And evidence brought forth to test this hypothesis probably has more to say about the likelihood of strategic bias than it does about measurement errors due to the hypothetical nature of the instrument.

Similarly in the next section CBS quote two sets of authors on the role of time in gathering information and learning about preferences. They then formulate the null hypothesis: the value expressed with little time for learning will be equal to the value expressed after the passage of time. But clearly what matters is not he passage of time alone, but whether that time is used to gather information about and experience with the new level of th environmental good. And the data reviewed by CBS do not shed much light on this question.

If the arguments offered here about unfamiliarity and learning are accepted, then it follows that any individual's response to a CVM question about a large change in the environmental good form the existing familiar level will include a potentially large random error component. But if these "hypothetical" errors are truly random with zero mean, then they will tend to cancel out over large samples. Thus with adequate sample size, sample mean responses may not be seriously inaccurate measures of the true mean values of the population.

Some authors have argued that there may be a systematic component to

the kind of hypothetical error I have been discussing here. For example, Bishop, Heberlein, and Kealy (1983) argued:

> "One resulting hypothesis worth future investigation is that people respond as they do to contingent markets because of uncertainty (presumably concerning their preferences). this may lead them to state answers which imply conservatively high requirements for compensations, amounts at which they are relatively certain they really would sell. They would even recognize the possbility that they might sell a lower amounts, but still give conservative answers in order to "play it safe." (p. 629)

Rendall, Hoehn, and Brookshire (1983, p. 643) reach similar conclusions on the basis of more formal analysis. The required compensation for the loss of an environmental good is that sum which enables the individual to remain at the initial utility level after the loss. That sum is found by solving the expenditure minimization problem for the initial utility level. if because of ignorance the individual does not find the expenditure minimizing solution, he will ask for higher compensation, thus overstating the willingness to accept compensation. A similar argument yields the conclusion that the stated willingness to pay for an increase in the environmental good will be less than the true value of willingness to pay. The argument is based on the assumption that individuals know their preferences will enough to identify alternative consumption bundles which yield the same initial level of utility but make mistakes in determining which of these bundles minimizes expenditure. But if individuals also lack accurate information on their preferences, they can make mistakes in attaching utility levels to different consumption bundles. Thus they may base willingness-to-pay responses on consumption bundles which turn out to yield either more than or less than the initial level of utility and thus state willingnesses to pay that are either less than or more than the true value. The Bishop, Heberlein, Kealy end Randall, Hoehn, Brookshire arguments are based on a more limited kind of ignorance. Ignorance that extends to the specific characteristics of one's preference orderings implies random rather than systematic errors in stated values.

E. ASSESSMENT

Chapter VI is perhaps the most interesting chapter on the report in that it is here that the authors confront the question of accuracy head on and discuss comparisons of CVM values with values derived from other methods. In this section I will offer some comments on their assessment and provide my own assessment of the CVM.

In reviewing the evidence form comparative studies, CBS make it clear that these comparisons are at best suggestive because of inaccuracies inherent in the TCM and HPM. Any quantitative estimate of the accuracy of the CVM requires that we know the true value being measured. Yet the HPM and TCM have errors that may be large, are not well understood, and are arguably of the same order of magnitude as those associated with the CVM. Their discussion of this point is a refreshing, perhaps chilling, reminder of the limitations of our empirical models. 4/

CBS conclude that if certain reference operating conditions are satisfied, the range of error associated with a CVM estimate of value is likely to be plus or minus 50 percent. This statement has a very ad hoc quality. I have some criticisms of the reasoning offered be CBS as to how they reached this estimate of accuracy. And I am not sure how it is meant to be interpreted. They do not distinguish between bias and random error in measurement. However their discussion on pages 190-191 appears to focus on biases. As I have argued above, not all of the kinds of bias they mention need to be present in a well-designed CVM study. Nor do all types of bias necesarily operate in the same direction and therefore decrease accuracy. Two biases of equal magnitude but opposite sign can offset each other resulting in an accurate measurement.

CBS appear to be making a statement about the accuracy of the aggregate value derived from a CVM study. Yet most of their argument deals with possible errors in individuals' bids. There is no discussion of the relationship between errors in individuals' bids (systematic or random) and errors in the aggreate value, or of the influence of sample size and aggregation technique on errors in aggregate value. The effect of random error in the measurement of individual values on the aggregate measure obviously depends upon sample size among other things.

Any quantitative assessment of the accuracy of the CVM must begin with the description of the CVM instrument to which it applies. The assessment should have two components. The first is a consideration of the likelihood of bias from each of the sources of bias discussed above and if possible an estimate of the likely magnitude and direction of each possible bias. The second is a consideration of the description of the environmental change being valued and of the respondents' familiarity with the environmental good and experience with changes in the environmental good over this range. If the CVM instrument is carefully designed to minimize the likelihood of various kinds of bias, and if the familiarity and experience criteria are satisfied (as for example in the Los Angeles air pollution study of Brookshire et al., 1982), then I would not be surprised if we could argue for accuracies substantially better than plus or minus 50 percent in the aggregate. however, ever if biases are minimized but the instrument calls for individuals to consider positions outside the range of experience and familiarity (as for example in the case of existence or preservation value fr unique environmental resources), then one cannot be so sure about the likely accuracy. This is because what is involved is the larger but, we

hope, random error in individual responses perhaps beings offset by large sample size.

To close this section, I would like to offer a somewhat more formal framework for the consideration of the question of accuracy. Let B* denote an individual's true bid or willingness to pay for an increase in the provision of the environmental good. Let B be the individual's response to a CVM question and assume that B is a random variable with a mean B'. The question of bias comes down to whether B' is greater than, equal to, or less than b*. The random component of measurement error is e = B - B', which has a zero mean. The analysis of the accuracy of the CVM response must focus on the magnitude of B' - B* and on the variance of e.

Consider the case of starting point bias. Assume for the moment that there are no other sources of bias and that for the individual e is identically equal to zero. Suppose that a set of identical individuals were asked CVM questions using one of the two approaches I suggested above for mitigating starting point bias in the aggregate mean bids. 5/ Although I haven't given the matter much thought, it seems possible to argue that the expected value of the mean bid is equal to B*. In other words, starting point bias in individual bids may be treated in such a way as to result in only random measurement error in the aggregate. It may be possible to develop similar arguments for the other sources of bias in individuals' responses.

Let us now assume that all bias problems have been successfully dealt with in the design of the CVM so that B' = B* for all of the identical individuals. Asking the CVM question of a sample of the population of identical individuals yields an estimate of B*. And of course, the accuracy of this estimate increases with the size of the sample. Very large variances in the error term in individual responses can be compensated for if the sample is large enough. It may be that the so-called problem of hypothetical bias is not that serious, at least if the error in hypothetical setting is really random. 6/

F. TWO MISCELLANEOUS COMMENTS

My first comment has to do with CBS's suggestion that the frequently observed large differences between willingness-to-accept-compensation questions and willingness-to-pay questions may be due to cognitive dissonance. CBS do not spell out their line of reasoning on this point, and I am not able to provide a convincing explanation based on what I understand about cognitive dissonance. If CBS have such an explanation in mind, it would contribute to our understanding of this puzzling empirical phenomenon if they were to make it explicit. Note that it is not sufficient for the thoery to predict willingness to accept being greater than willingness to pay. We already have such a theory based on income effects. to be helpful, the theory should predict potentially large differences.

My second comment concerns the inferences that CBS draw from experiments with the second price auction for the design of CVM instruments. They say:

> "Individuals must be placed in an "all or nothing" situation in the questionnaire where no strategic holding back can help them. ... Secondly, an iterative option framework is suggested. Because of the reported demand revelation "learning period" associated with the second price auction, individuals also should be placed in a survey situation which provides them with tentative information about allocations before results are finalized." (p.90)

And in footnote 6 they go on to say:

> “That is, provide the individuals with more than a one-shot survey. Let them answer a survey, report the tentative results of that survey back to them, let them adjust their answers, report the new tentative results, and so forth until an unannounced stopping time. At the stopping time allow the final results to take effect" (p. 102-A)

I have two comments concerning this suggestion. First, the second price auction provides a rule for determining the price of the actual transaction. Its purpose is to eliminate the incentives for strategic behavior on the part of bidders. But in a CVM survey, there is no actual transaction and, we hope, no incentive for strategic behavior. Thus no purpose is served by presenting survey respondents with a second price rule. In fact, this further complicates the survey instrument and may lead to confusion on the part of respondents.

My second comment concerns their proposal to report back information on the aggregate bids and carry out further iterations. This procedure proved useful in experimental settings where the end result was an actual transaction. CBS argue that this procedure helped participants to learn about the incentive compatibility feature of the second price auction where actual transactions are to take place. But the iteration procedure probably does not help individuals to learn more about an unfamiliar region of their preference ordering. Thus the iterative procedure does not seem likely to contribute to a reduction in the random measurement error associated with the hypothetical nature of the CVM survey.

G. CONCLUSIONS

I will conclude by offering some summary comments about the Assessment offered by CBS and then offering my own assessment. On the positive side, I think this Assessment makes a substantial contribution in the following respects: (1) its emphasis on the question of accuracy of responses; (2) the effort to base the Assessment on the formulation and testing of hypotheses concerning such things as biases and sources of error; (3) the introduction of the notion of familiarity with the environmental good and experience with changes in its quantity as important conditions for extracting accurate measures of value.

On the other hand, the CVM technique for eliciting monetary values from respondents represents only one member of a family of contingent choice techniques. It would have been useful to consider the extent to which all of the members of this family suffer from similar problems due to their hypothetical nature as well as to consider the relative strength and weaknesses of these different approaches to estimating values. Second, the Assessment should have incorporated a more precise definition of reference accuracy and an analysis of the separate roles of bias and random error in determining the degree of accuracy of any specific contingent choice technique. Finally, it would have been helpful to integrate the concepts of familiarity and experience into their discussion of hypothetical responses and their efforts to test hypotheses in Chapter V.

My comments on the CVM itself are somewhat encouraging in one respect. that is, at least some of the bias problems appear to be manageable; and if measurement errors due to the hypothetical nature of the instrument are random and not too large, then larger sample size is a potential means of coping with them. However, there is a negative side of this assessment. On the basis of the familiarity and experience arguments, it appears that the CVM is likely to work best for those kinds of problems where we need it least; that is, where respondents' experience with changes in the level of the environmental good have left a record of trade-offs, substitutions, and so forth, which can be the basis of econometric estimates of value. But for those problems for which we need something like the CVM most, that is, where individuals have little or no experience with different levels of the environmental good, CVM appears to be least reliable.

ENDNOTES

Chapter X.

1) Thayer's comparison of values obtained by the CVM and site substitution models is a comparison between techniques which belong to the same family of contingent choice or hypothetical valuation approaches. Thus the comparison should not be construed as a test for hypothetical bias (CBS, p. 173).

2) The equivalent measure of value can be defined in a similar manner. Some CVM studies have sought to obtain equivalent measures in the form of willingness to pay to avoid the loss of an environmental good. See, for example, Brookshire, Ives, and Schulze (1976) and Desvousges, Smith, and Megivney (1983).

3) Alternatively Carson and Mitchell (1984, p. 16) suggest using one of the available techniques for imputing missing willingness-to-pay values on the basis of the remaining sample data.

4) Not all estimates of the benefits of environmental improvements are subject to inaccuracies of this magnitude. For example, if an improvement in air quality in a small region leads to an increase in the output of an agricultural commodity without significant input or crop substitution effects or impact on market price, then the observed increase in output can be combined with a presumably accurately measured market price to yield a reasonably accurate measure of the benefits of increased output. The problems of estimation arise when there are significant price effects and behavioral responses which must be modeled and quantified to produce defensible benefit estimates.

5) That is, either dividing the group equally and employing an appropriately set low starting point with one group, etc., or using the "bracket and half" technique with randomly chosen staring points.

6) For example suppose that we interpret CBS's estimate of a plus of minus 50 percent error to refer to the individual response error and (assuming that e is normally distributed) to mean that the interval of B* plus or minus two standard deviation is .5B* - 1.5B*. Alternatively the probability is approximately .95 that B will be in this interval. A sample of 16 identical individuals is sufficient to reduce the error of the sample mean as an estimate of B* to $\pm$ 12½ percent. Similarly , if the error in the individual responses is plus or minus 100 percent, a sample of 100 individuals yields an error of plus or minus 10 percent.

XI. TO KEEP OR TOSS THE CONTINGENT VALUATION METHOD

BY
V. Kerry Smith
Centennial Professor of Economics
Vanderbilt University*

A. INTRODUCTION

In concluding his essay on the rhetoric of economics, McCloskey (1983) dicussed the role of surveys under a subheading "Better Science," presumably intended as an admonition to the economics profession. He observed that:

> "Economists are so impressed by the confusions that might possibly result from questionnaires that they abandon them entirely, in favor Of the confusions resulting from external observation. They are unthinkingly committed to the notion that only the externally observable behavior of economic actors is admissible evidence in arguments concerning economics." (p. 514)

He continued this discussion, questioning such views by acknowledging that:

> "Foolish inquiries into motives and foolish use of human informants will produce nonsense. But this is also true of foolish use of the evidence more commonly admitted into the economist's study." (p. 514)

Of course, these comments should not be interpreted as an endorsement for the contingent valuation method. Rather they represent a call for a more open attitude in judging the sources of information used in evaluating (or implementing) economic models. At the same time, however, they do present a reasonably accurate summary of the attitudes of a majority of economists. While there has been somewhat more acceptance of the potential usefulness of survey information associated with individuals' or firms' attitudes or plans, these are always regarded as less desirable sources of information relative to "hard" statistical data or the predictions of econometric models based on those data. 1/

Unfortunately, environmental economics encounters a wide range of resource allocation decisions wherein we would not expect, because of the nature of the resources themselves, the market interactions of economic agents to reveal information which would assist with these decisions. Many, if not most, environmental resources exchange outside markets; they exhibit some of the features of public goods; and they are not easily measured or translated into a quantitative scale. For example, good air quality implies an absence of air pollutants. Thus, we might consider measuring it by using this relationship and records on the ambient concentrations of pollutants. However, these technical measures do not necessarily translate readily into eiher the household's perceived air quality or the features of pollution

which impair health or the aesthetic dimensions of the environment. 2/ As a result of all these limitations, the empirical practice of environmental economics has come increasingly to rely on the use of direct interviews to obtain information on individuals' valuations of environmental resources.

Increased interest in and requirements for measures of the benefits associated with changes in one or more aspects of environmental resources have focused attention on the use of the direct interview or contingent valuation method as a basis for deriving such estimates. The objective of this paper is to use the recent comprehensive review and evaluation of the contingent valuation method by Cummings, Brookshire, and Schulze (1984) as the basis for an independent appraisal of the method and, with it, a commentary on these authors' judgments.

Cummings et al. have provided a thorough review of the conceptual and empirical issues associated with contingent valuation methods (CVM). Their study has integrated a large and diverse set of CVM studies and attempted to extract from them a summary of what this work has determined concerning the performance and viability of the contingent valuation method. One interpretation of the authors' bottom line (or reference operating conditions) would suggest that: CVM can be expected to perform best for commodities where we would be least likely to want to use it. That is, respondents should be familiar with the commodity, have choice experience assocated with its consumption, and be relatively certain about the conditions of availability posed in any CVM valuation question. In these circumstances there are often other methods for estimating individuals' valuatons of environmental amenities (see Freeman, 1979a). Indeed, it is the presence of these other methods for such cases that has provided the opportunity to perform comparative analyses of the benefit estimates derived using CVM in relation to another indirect method (i.e., one based on the observable actions of households). These comparative analyses have, in turn, led to the definition of the Cummings et al. reference operating conditions. When we relax one or more of the reference operating conditions, the authors suggest that the performance of CVM cannot be easily judged. This conclusion is not surprising because there does not exist a basis for a comparative analysis of estimates from different methods in these cases.

Rather than cover the same groundwork developed in the Cummings et al. analysis, we will approach the evaluation of CVM from a somewhat different perspective. Assume that the objective of CVM research is the estimation of individuals' valuation of changes in specific environmental amenities (so that each type is consistently reflected in these valuations). Given this goal, it should be acknowledged at the outset that we will never know how well CVM or any other method performs in estimating their "true" valuations. Consequently, to evaluate these methods we have two choices. We can formulate a model that describes the consumer's decision process, including the valuation of the relevant amenities, examine within the context of that model how CVM's responses would be made, and compare the model's prediction of those responses with the model's true valuations. Alternatively, we can attempt to devise an experimental setting that would mimic the essential elements present in a real-world CVM application (tailored to the limits of the experimental setting), collect data on responses, and evaluate CVM in comparaison with what was expected from the

experimental design.

Both approaches require assumptions to use their respective findings in evaluating CVM in a real-world context. For the first it is a matter of the correspondence between the model of consumer behavior and its representation of individuals' responses to CVM in comparison to reality. Not all maintained hypotheses can be tested in the absence of knowledge of individuals' true valuations. In the second, a similar issue arises in the authenticity of the experiments' description of the actual decision process. Experiments necessarily require simplifications (as do models). Relating the findings from each approach to the performance of what is evaluated in the real world involves gauging the importance of these simplifications. In short, professional judgment plays a significant role in either of these exercises. As a consequence it seems reasonable to begin an evaluation of CVM with an inquiry into the realization of these judgements in the appraisal of other economic data bases. That is, in what follows, we consider a selected set of surveys, involving both households and firms, and examine the attributes of some of the questions posed in these surveys. Based on this partial review, it appears that in many cases our objective data are based on questions that require judgements, responses that may be subject to strategic biases, and valuation responses under hypothetical circumstances. Indeed, they are subject to many of the problems discussed as if they were exclusively associated with CVM data. Moreover, some of these "offenders" (i.e., cases where the effects of these sources of bias may be important) involve the data that have been used in several of the indirect approaches to benefit estimation. Following this review, Section C discusses in more detail the attributes of the questions that are asked and how these characteristics appear to affect our willingness to accept individuals' (or firms') responses as objective data. While there are a number of considerations associated with what Medoff and Abraham (1979) describe as "having contact with units of observation" (see Note #1), the most important stumbling block to the CVM approach appears to be the combination of a hypothetical question and changes in the resources that are outside the range of an individual's experience. Consequently, Section D discusses the implications of the arguments against using responses to hypothetical situations as indicative of consumers' valuation should these situations in fact be realized.

The last section considers what this perspective on CVM implies for the use of its results and for further research. An appendix clarifies some inaccuracies (in this author's opinion) in the Cummings et al. summary Of the research.

B. NON-CVM DATA: HOW OBJECTIVE ARE THEY?

Table 11.1 summarizes a sample of data sets that are used in a variety of economic models. While many have a direct relation to empirical studies in environmental economics, they are not exclusively so. In each case, one of the the uses of the data, the name of the survey, a variable observed, the questions used to derive it, and a judgemental evaluation of the response are reported.

There are several aspects of the Table which are relevant to an evaluation of CVM. First, and perhaps most interesting, responses to hypothetical questions play a prominent role in two of these cases. The hedonic property value model, usually regarded as the most promising indirect, market-based alternative to CVM has often been based on data from either the Annual Housing Survey or the Census of Population. 3/ Both data sources report, for owner-occupied units, the owners' appraisal of their selling prices if they were to sell their homes, not the market prices. Thus, hedonic models based on these data reflect the owners' perception of the prices they would realize and not necessarily the equilibrium locus as hypothesized. These individuals are being asked a hypothetical quesiton and it should clearly be recognized as such. Of course, it may be reasonable to assume that the respondents form their perceptions of the relevant market price based on past sales in their neighborhoods. Nonetheless, this is not necessarily a good proxy for actual price. It will depend on the number of homes selling in their neighborhood, as well as on each individual's ability to translate these sale prices into a corresponding estimate of the price of his (or her) home. There does not appear to have been a comparison for specific cities of the results that would have been derived using the Survey or the Census in comparison to the use of the actual sales and their implied hedonic price function. Therefore, it is difficult to judge the implications of the use of these hypothetical data.

Another example with hypothetical responses playing a tangible role in the development of "hard" or objective data arises in one of the constituents of the CPI. In January 1983 the CPI changed its treatment of the components of the cost of shelter. Under the old method, this cost was measured based on changes in the cost of five items -- home purchase, contracted mortgage interest rates, property taxes, property insurance, and maintenance and repair. The new approach attempts to measure the change in the cost of obtaining, in the rental market, housing services equivalent to the rental home. These are measured with actual rents. However, the weights used to reflect their contribution are based on a question in the 1972-73 Consumer Expenditure Survey asking households for how much they think their home would rent. 4/ This is a hypothetical question which may well be more difficult for households than to gauge the selling price for a home, especially since their knowledge of the relevant rental market may be quite limited.

Secondly, there are incentives for strategic responses even in the questions reporting so-called "hard" data. One of the more controversial of these concerns the reporting of employment status for young men. 5/ Discrepancies in the implied unemployment rates based on the Current Population survey (CPS) and the National Longitudinal Survey (NLS) of Young Men (see Freeman and Medoff, 1982) have led to several studies to

TABLE 11.1: SELECTED SAMPLE OF SOURCES FOR NON-CVM ECONOMIC DATA

OBJECTIVE	NAME OF SURVEY	VARIABLE(S) MEASURED	QUESTION	EVALUATION OF RESPONSE
1. Benefit Estimation				
Hedonic property value	Annual Housing Survey and Census of Population	Measure of the price of housing unit[a]	What is the value of this property, that is, how much do you think this property would sell for if it were for sale?[b]	Asks for estimate of value under hypothetical situation
Hedonic property value	Annual Housing Survey	Measure of perceived neighborhood amenities	Do you have satisfactory ____? (Public transportation, schools, police protection, neighborhood shopping, outdoor recreation, hospitals and health clinics)	Asks for judgment as to satisfactory
2. Benefit Estimation[c]				
Recreation participation models	Heritage Conservation and Recreation Service General Population Survey	Measure of recreation use	I am going to read you a list of outdoor recreation activities. For each activity please tell me if you have participated in this activity during the past 12 months. (30 activities considered, number of times requested or desire to)	Asks for extensive recall of detailed activities

(continued)

Table 11.1, continued

OBJECTIVE	NAME OF SURVEY	VARIABLE(S) MEASURED	QUESTION	EVALUATION OF RESPONSE
Recreation participation models	Heritage Conservation and Recreation Service General Population Survey	Demand response	If the price of gasoline doubled within the next 6 months, would this be likely to limit or curtail the number of trips you might take by automobile for outdoor recreation activities?	Asks for hypothetical demand response
3. Benefit Estimation				
Damage function	Health Information Survey	Measures of morbidity effect	Not counting the days in bed, lost from work and lost from school, were there any (other) days during the past 2 weeks that _____ cut down on because of illness or injury?[d]	Asks for recall and judgment of usual
4. Benefit Estimation				
Hedonic wage models	Current Population Survey (and others)	Measure of wage rate	How much does _____ usually earn per week at this job before deductions? How many hours per week does _____ usually work at this job?	Asks for recall for different individuals and judgment as to usual

(continued)

Table 11.1, continued

OBJECTIVE	NAME OF SURVEY	VARIABLE(S) MEASURED	QUESTION	EVALUATION OF RESPONSE
5. Price Index Measures				
CPI weights	1972-73 Consumer Expenditure Survey	Weight for owner's equivalent rent	If you were to rent out your home today, how much do you think it would rent for monthly, unfurnished, and without utilities?	Asks for judgment in hypothetical circumstance
6. Employment and Unemployment Measures				
Determinants of employment	Current Population Survey	Measure of unemployment	Has _____ been looking for work during the past 4 weeks? Does _____ intend to look for work of any kind in the next 12 months?	Asks for judgment "on looking for" by individual other than respondent, and of intentions in the future
7. Household production[e]	Income Dynamics Panel	Measure of value of repairs	During (year), did you (or your family) do any of your own repair work on your cars? About how much do you think you saved doing this last year? Was it about 25, 50, 100, 200 or what?	Asks for recall and judgment of market price

(continued)

Table 11.1, continued

OBJECTIVE	NAME OF SURVEY	VARIABLE(S) MEASURED	QUESTION	EVALUATION OF RESPONSE
8. Cost estimates	Survey on Pollution Abatement Costs and Expenditures	Abatement cost recovered	Report your best estimate of the value of materials or energy reclaimed (costs recovered) through pollution abatement activities and either reused in production or sold by form of pollution abated. (Exclude the value of items if they would have been recovered, sold, or reused in production in the absence of any pollution control regulations.)	Asks for cost saving allocation; has strategic incentives; judgment on no-regulation case
Cost estimates	Survey on Pollution Abatement Costs and Expenditures	Annual operating costs for pollution abatement	Report the annual operating costs and expenses for pollution abatement in (year). Distribute total operating and maintenance cost in terms of percent by form of pollution abated (air, water, solid)	Asks for cost allocation; has strategic incentives

Footnotes:

[a]This question applies to owner-occupied housing units. A comparable question is used to acquire rents for occupied rental units.

[b]Emphasis added.

(continued)

Table 11.1, continued

[c]Similar questions asked in the site specific Federal Estate Survey along with recall of past trips to an area and number of hours of travel.

[d]This question is asked after similar inquiries on days in bed, lost from work, and lost from school. It is asked of one person for each member of the household.

[e]This is one small aspect of the panel's questionnaire which deals with nearly all aspects of a households' activities, including information on the attitudes and feelings of its members.

Sources:

Annual Housing Survey: 1980, Current Housing Reports, H-170-80-1, February 1984, U.S. Department of Commerce, Bureau of the Census, U.S. Department of Housing and Urban Development.

Consumer Expenditure Survey: Diary Survey 1980-81, Bulletin 2173, September 1983, U.S. Department of Labor, Bureau of Labor Statistics.

1980 Census of Population and Housing, Users Guide, Part A, U.S. Department of Commerce, Bureau of the Census.

Bureau of Labor Statistics, Handbook of Methods, Vol. II, The Consumer Price Index, Bulletin 2134-2, U.S. Department of Labor, Bureau of Labor Statistics, April 1984.

Bureau of the Census, U.S. Department of Commerce, Current Industrial Reports, Pollution Abatement Costs and Expenditures, 1982, MA-200(82-1, February 1984.

Survey Research Center, A Panel of Income Dynamics, Vol. 1, (Institute for Social Research, University of Michigan, 1972).

Paul R. Portney and John Mullahy, Ambient Ozone and Human Health: An Epidemiological Analysis, Vol. II, Draft Report to Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, September 1983 (source of questionnaire for 1979 Health Interview Survey).

Heritage Conservation and Recreation Service, U.S. Department of the Interior, The Third Nationwide Outdoor Recreation Plan, Appendix II, Survey Technical Report 2, December 1979.

investigate reasons for differences in responses based on essentially the same questions. It should also be noted, as the entry in Table 11.1 for the CPS indicates, these questions impose additional requirements on respondents by calling for an interpretation of "looking for" work and an appraisal of an individual's future intentions. Both issues are reported by proxy respondents for youths with the CPS survey and by the youths themselves with the NLS. Freeman and Medoff (1982) report some evidence that the differences in responses used to infer unemployment rates may be biased at least partially because the proxy respondent's self esteem (in the CPS) was affected by the answers given.

The responses by firms to questions on pollution abatement costs also provide a case where strategic responses would seem likely to be a factor in interpreting the quality of these data. To date, however, there appears to be increasing use of these data without appreciable concern for these biases. 6/

A third area involves requests for "sensitive" information. These requests have long been recognized to offer the potential for biased responses. Questions involving income and wage information are examples. The latter has also served in indirect benefit estimates (hedonic wage models). While recent estimates of the magnitude of the differences between means of self-reported and employer-reported wage rates seem fairly large (i.e., 4.8%) and are significantly different from zero, 7/ Mellow and Sider's (1983) overall results indicate that "... the estimated structure of the wage determination process is essentially independent of the source of information." (p. 342)

There are further examples in Table 11.1. However, these three classes of problems are sufficient to draw attention to the potential for significant limitations with many (if not all) objective data sources for economic analysis. Only artificial data (i.e., data generated from a controlled model) are perfect. This is hardly surprising and not the point.

When any data are derived from surveys we can expect they will be subject to limitations. Nonetheless, with the major surveys similar to those identified in Table 11.1, these limitations have been accepted as tolerable. Results derived from most of these data sources are routinely accepted by the relevant subset of the economics profession as plausible -- not as the last word on any subject, but rather they are judged to be worthy of consideration and review, as constituents to a body of developing empirical evidence on a particular subject. In effect, they have passed an implicit standard of tolerance for the quality of data. BY contrast, data from CVM experiments appear to fall below this standard in the judgement of the majority of economists. Consequently, one approach to understanding the potential limitations with the contingent valuation method is to examine the reasons for these revealed preferences of economists. That is, we must consider what attributes of CVM prevent its data from passing the professional "muster."

Before addressing this issue, however, it is important to recognize, as Mitchell and Carson have observed in their recent review of the Cummings et al. appraisal (Appendix to Chapter XIII) -- not all CVM studies have been of equal quality. Not only have the sample sizes been quite small in some cases, but quality of the questionnaires used to elicit responses to complex questions has also been diverse. This is to be

expected since the development of CVM has been a learning process. Thus it should be acknowledged that past quesionnaires have introduced confusion in what was elicited and my not indicate the prospective performance of the method with appropriate attention to questionnaire design. The debate over the interpretation of Greenley, Walsh, and Young's (1981) estimates of option value (see Mitchell and Carson, forthcoming) is but one example where what was communicated to respondents is at issue since it provides the only basis for the results.

Unfortunately, the Cummings et al. review seems to treat all CVM studies as if they conveyed equal information on the properties of the method. Clearly, they do not. It is, of course, difficult to judge on the basis of the published summaries of such studies where these limitations might be. Since this issue has implications for future research, it will be discussed in the last section of this paper. At this point, it is important to note that the available CVM estimates reflect both a learning process in the use of questionnaires (as economists discovered the survey research relevant to eliciting value information) and the inherent properties of the approach as a basis for valuation information. Separation of these two influences inevitably involves judgement. This judgement is reflected in the contrast between the Carson-Mitchell (1984) appraisal of the sources of error in CVM and that of Cummings et al. Nonetheless, even with these problems, there do appear to be features of what CVM asks that can be distinguished from what is elicited in the surveys that are judged "acceptable" by most economists.

C. TASKS REQUESTED OF SURVEY RESPONDENTS

Surveys request individuals to undertake a number of different types of response tasks. The list below attempts to classify and describe each type of task. They have been ordered according to what appears to be (based on an admittedly limited reading) the profession's perception of the likely accuracy of the responses.

(a) recall: to remember patterns of behavior over some past time period (often in detail). This task can include requests for information on the actions of the individual or of members of the household. It can extend as long as a year and require a time-sequenced report, either through an ongoing diary or an ex post report;

(b) partitioning: to assign a portion of time or expenditures to engage in certain activities or meet particular objectives. A detailed accounting of the types of recreational activities undertaken and the days spent at each is an example of this task from Table 11.1;

(c) judgement of a state: to appraise a condition based on a described set of criteria, e.g., seeking work or evaluation of health status;

(d) truthful response on sensitive information: to report sensitive financial or personal information that may be factual but is regarded as confidential by the individual, e.g., income or assets;

(e) evaluation of attitudes: to evaluate sentiments and feelings with regard to an issue or condition;

(f) projected responses to hypothetical circumstances: to describe actions under proposed conditions that have not occurred, e.g., what would a person do if some action took place; or to judge what he or she perceives another individual or institution would do if an action took place.

The first three tasks seem relatively uncontroversial. While there is some tendency to question aspects of information derived from these types of inquiries, with our discussion of concern over available measures of the employment status of young men as one example, these issues have not led to the dismissal of the data involved. There is a large literature in survey research on the question of sensitive information. Income questions are always at the end of a questionnaire. The income supplement to the CPS, for example, is asked of the group rotating out of the sample, not of the individuals expected to continue to be a part of the survey whose future participation and responses are valued. Nonetheless, when responses are given, they are routinely accepted for subsequent economic analysis.

For the last two categories, however, economists are at best skeptical of the merits of the information. Cummings et al. acknowledge the mistrust of attitudinal data. Both their volatile nature and the difficulty in developing standards for gauging the comparability of these responses across individuals has limited their acceptance in economics. At the same level of acceptability as attitudinal information, or perhaps below, come the tasks involving hypothetical questions. This is why CVM is faced with justifying the plausibility of its information.

However, our brief overview of some established survey data bases indicates that they also involve responses to hypothetical questions. Yet, in these cases, the concerns that economists express with CVM do not appear to have been raised. Why? The answer seems to be fundamental differences

in the hypothetical tasks requested. Markets do exist for the commodities involved and it is assumed that the individual is fully aware of them. Consequently, under the most favorable interpretation, the responses that are requested could be considered as asking the individual to match his (or her) commodity with the relevant existing market and report the current price. Of course, the nature of the markets for heterogeneous commodities, such as housing, are not completely consistent with this view. Moreover, each individual's knowledge of these markets can be expected to vary; and this requested matching process will be affected by the individual's perception of his (or her) home. Nonetheless, the nature of what is asked is fundamentally different. It is not to search one's preferences, recognize financial constraints, and respond with a bid. Rather it is to report what each individual perceives the market would yield as a price or a rent for an existing commodity.

It appears that Cummings et al.'s reference operating conditions impose a similar requirement on CVM. That is, under their ROC, individuals must have had the ability to obtain "choice experience with respect to consumption levels of the commodity." This implies that there is some mechanism available to individuals to enable them to select the different levels of the resource involved. If there are not formal markets, then we must ask what the mechanism is. If it leads to the equivalent of an implicit market, then we must assume that choice experience is the equivalent of knowledge of the features of the implicit market. Indeed, Cummings et al. state as much in their closing arguments, observing that:

> '... The state of the arts is one wherein we can simply say that evidence exists which supports the proposition that indirect market exprience with a commodity may serve to satisfy the ROC's: when the environmental good is a distinct attribute of a market-related good (water quality in a time/travel cost recreation trip or air quality as an attribute of housing locations/costs), experience/familiarity with the market good seemingly spills over to the individual's ability to value the attribute." (p. 207)

Consequently, reference operating conditions amount to a requirement that we accept CVM studies only where they involve hypothetical questions comparable to those in existing surveys -- asking for implicit market outcomes for hypotetical changes. This is not the same as asking an individual's bid for a commodity that is not exchanged.

Consequently, the most important limitation to the acceptance of CVM appears to be its use in eliciting an individual's response to a hypothetical situation. Responses that involve individual judgements as to the nature of market outcome (either formal or implicit) in response to a hypothetical change are not viewed with the same degree of skepticism. Therefore, to evaluate the prospects of CVM we must consider why the responses to these questions are viewed as unreliable and determine if there is existing or new research which might resolve the issues involved.

D. THE PROBLEMS WITH HYPOTHETICAL QUESTIONS

The principal problems with hypothetical questions concerning an individual's behavior can be summarized using three questions:

(1) Will each respondent really take the decision circumstances seriously, since there are no tangible incentives to do so?
(2) Is an individual capable of processing the information involved in what is often a completely new (or at least an unfamiliar) set of conditions, and responding with his or her actual valuation, even though this value would ordinarily be derived after time for consideration?
(3) Does an individual's response require repeated experience to form an appraisal of the valuation of the hypothetical question?

The first and third questions are components of Feenberg and Mills' (1980) critique of the survey approach as a basic source of valuation information. While all three are identified in Cummings et al. discussion, these authors do not explore their implications for other methods for benefit estimation. That is, indirect methods which are based on households' observed behavior would also be affected by the decision frameworks implied by questions (2) and (3). All indirect approaches assume the individual has complete information on the available commodities (including those whose purchase is tied to the receipt of an environmental resource). If repeated experience is necessary to form a judgement on the features of the resource and to value it, then the role of experience must also be reflected in the models used to derive indirect measures of households' valuations of environmental resources. Of course, these questions are not independent. Repeated experience provides information that may assist in the decision process described (i.e., question (2)). None of the existing indirect benefit measures reflect this type of decision process. Thus, if this view describes behavior then all of the indirect methods will be biased in an unknown way.

Both approaches to estimating individual's valuations for nonmarketed commodities involve hypothetical conditions. To use either approach requires a judgement of the correspondence between their predictions (or responses) and actual behavior. For the indirect methods we formulate a hypothetical description of an individual's behavior in the presence of a specified characterization of what is known and what constrains decisions. This framework is then used to evaluate actual decisions as if they were guided by it. The direct or contingent valuation approach formulates hypothetical circumstances and asks what an individual's behavior would be. Neither escapes the hypothetical. Consequently, criticisms that are based on a belief that individual decision processes are too complex to be adequately determined from one-time hypothetical questions will also be relevant to the indirect methods.

Of course, what is important is by how much is each approach affected by its respective assumptions. Cummings et al. results suggest we don't know the answers for the contingent valuation method. However, it seems the same conclusion would be drawn for the indirect approaches. Few economists would contend that housing markets behave in accordance with the hedonic model -- assuming that we can exactly measure an equilibrium price structure

with home sales within any period. However, there does appear to be a reasonably wide consensus that, despite the errors introduced by departures from equilibrium, the estimates of the marginal willingness to pay for site attributes are usable. That is, it is tacitly assumed that these errors are not sufficiently large to invalidate the practice. In fact, there has been no appraisal of the extent to which the model's assumptions affect its performance. Judgemental evaluations of Maler (1977) and Freeman (1979b) are at opposite extremes in terms of their respective interpretations of the importance of the model's assumptions. Thus, if one accepts these criticisms of the contingent valuation method, it is unlikely that comparative analyses of CVM to indirect approaches, whether hedonic property value or travel cost, will resolve matters.

What is needed is an evaluation of the models as they have been asked to perform. For example, with the hedonic property value model we might ask:

(a) Does an equilibrium matching of buyers and sellers under real-world conditions lead to a smooth continuous price function?
(b) Is the specification for the equilibrium price function derived under the conventional fitting criteria of econometrics likely to provide accurate estimates of the marginal valuations of site attributes, such as environmental quality?
(c) Is the mechanism an individual uses to form perceptions of site characteristics (or diversity in mechanisms across individuals) important to the viability of the method?
(d) Can these marginal willingness-to-pay estimates be used to derive an individual's inverse demand for a site attribute?

The literature abounds with analytical answers to parts of these questions, but none are designed to comprehensively evaluate the methods under conditions that resemble the real world.

Equally important, we do not have a model of how individuals will respond to CVM questions. Hoehn and Randall (1984) have suggested that we can identify the direction of the errors by simply considering the Optimal strategies for participants within a simple model of their decision process. Their model identifies two key incentives to the character of participants' responses: judgements as to how participation is likely to influence a policy designed to increase the environmental amenity of interest; and judgements as to the level of disposable income if the policy is undertaken. Both rely on individuals acting strategically in their responses -- in effect taking the process seriously. Thus, while the Hoehn-Randall framework is an interesting beginning in the modeling of individuals' response to CVM, it does not address the fundamental issue -- how will individuals behave when their stake in the process is not clear? Some researchers have argued truth-telling is the simplest response. Others follow Feenberg and Mills indicating that they will be more likely to provide attitudes that will vary with whatever happens to be the most recent stimuli or information influencing these attitudes.

At this point there can be no answer to this issue until there is a model of the process itself. Moreover, there is unlikely to be a model forthcoming until those economists involved in CVM perform research on how individuals respond to these types of questions -- in effect, attempt to

understand what will guide individuals' responses to questions eliciting their valuations of hypothetical changes in nonmarketed resources. It should be acknowledged that economists have not had experience in this type of research. 9/ Moreover, there is no assurance that it will lead to sufficient information to permit the response process to be understood and modeled. There are, however, companion research efforts that with efforts to model responses to CVM should enhance our ability to judge CVM. They include:

(a) Evaluations of the Indirect Methods

Comparisons of indirect and CVM estimates are largely useless unless we can bound the nature of the errors associated with the indirect estimates. Evaluation of the performance of indirect methods under something resembling real-world conditions is essential to interpreting these findings. While such an evaluation will not establish results for CVM that would be relevant to its application under conditions without an implicit market, it can help to answer whether individuals will take CVM questions seriously in the absence of clear incentives or consequences for their behavior.

(b) Evaluate Infrequent and New Commodity Decisions

There is no reason why the issues associated with learning about the nature of a new commodity or judging how to interpret behavioral decisions with infrequently purchased goods could not be investigated for market commodities. What type of information is acquired? What are the roles of service and maintenance patterns, price, etc.? The analysis should provide empirical information on these issues that would be relevant to the interpretation of CVM in circumstances that involve completely new resources, one-time or very infrequent decisions, etc.

(c) Experiment with CVM Formats

As Cummings et al. acknowledge, laboratory experiments provide an opportunity to understand some elements of the performance of CVM. They can never provide the answers to all CVM questions because they also require assumptions to transfer their findings to real-world circumstances. For questions involving tie evaluation of institutional structures they can be invaluable. In understanding how individuals respond to hypothetical changes in an environmental resource, their value is more limited because the experiments require control, and with it simplification.

E. THE BOTTOM LINE

The objective of the Cummings et al. summary and analysis of the contingent valuation method to benefit estimation was to take stock of what has been accomplished and evaluate whether, despite most economists' skepticism concerning the method, its continued use can be justified in benefits research. In effect, can we hope for acceptance of CVM research results more generally by professional economists? These authors' conclusion recognizes that the only standard available from current research is itself an estimate of the unknown "true" value of an individual's valuation. 10/ Consequently, Cummings et al. must argue that the standard used in these comparisons has some level of accuracy -- i.e., it includes the true value in a plus or minus 50 percent confidence interval. With this assumption, then, the authors argue that CVM estimates derived from studies satisfying their reference operating conditions will lie within plus or minus 50 percent of the standard (i.e., the indirect estimate). Of course, there are an infinite number of ways that a CVM confidence interval could include the indirect estimate without having a comparable likelihood of including the true value. 11/ Their summary is a valiant attempt to use the available information to judge CVM. Unfortunately, it does not establish a confidence interval for the CVM approach. At this stage it cannot, without acceptance as a maintained hypothesis that individuals will attempt to report their true values and therefore the variation observed across individuals (after taking account of socio-economic characteristics), can be treated as a random error due to each individual's differential understanding of the full implications of what is asked.

Indeed, there are several general statements that can be made independent of the Cummings et al. appraisal concerning CVM.

(1) There has been no research designed to systematically evaluate CVM for benefit estimation. Moreover, we do not have the information available to develop a confidence interval for indirect benefit estimates applied under the conditions in which they are applied. Their assumptions are not satisfied and most economists recognize these failures. We do not know how much these violations in assumed conditions affect the performance of the estimates. The Cummings et al. reference accuracy for the indirect method is their judgemental interval estimate. What is the likelihood the true value will fall in this interval? We cannot answer that question. Indeed, on an analytical basis we may never be able to do so.

However, we can use our models to gauge the sensitivity of results to the assumptions most likely to be violated. This would seem a necessary first step in evaluating the available comparative evidence. Until we know how good the indirect methods are, it will be impossible to judge the meaning of proximity of point estimates from CVM and a particular indirect approach.

(2) One reason why there has been diversity among CVM researchers in their judgements as to its performance is the use they intend for the benefit estimates. In effect, one must ask how will the CVM estimates be used. We may be able to tolerate low levels of accuracy for some purposes. It appears that those evaluating CVM have quite different end uses in mind. The old adage -- "good enough for government work" -- may well be literally relevant in some applications of CVM estimates. Not all benefit-cost

analyses will require CVM estimates with the Same accuracy. A wide range of estimates may still permit a yes/no decision to be made. This was Hoehn and Randall's (1983) point some time ago.

By contrast, however, tests of specific hypotheses or indeed some benefit-cost decisions may hinge on the accuracy of the estimates of individual valuation. These end uses and their implied standards should be identified. CVM may prove acceptable in some cases and not others. We cannot hope to provide this type of answer if the questions fail to recognize the implications of the potential differences in the uses of CVM results for any evaluation of the methodology.

(3) At present the evaluation of CVM results is exceptionally difficult because of the lack of uniformity in reporting information. Broad professional acceptance of CVM results requires clear and comprehensive reporting of all the details of the survey. The estimates are only as good as each individual respondent's understanding of what is asked. External reviewers cannot hope to be aware of all of the details of each application. A uniform reporting system with the assurance of backup detailed information would facilitate the evaluation of the influence of questionnaire and survey design on the results.

There has been no research designed exclusively to evaluate CVM. Rather studies have been conducted to serve multiple objectives. In such a Setting it is essential to have full information on these design issues in order to gauge the plausibility of the CVM estimates.

The bottom line on CVM is not what Cummings et al. suggest. In this author's judgement we can draw no conclusion on its accuracy based on what we know from research to date. After over a decade's expeience with CVM, this is certainly not a satisfying conclusion, especially given the volume of research resources currently involved in using it for some valuation objective. However, this judgement must also be considered in the context of what we really know about other methods for benefit measurement. There is no more reason for being confident of the estimates derived from indirect benefit methods. The degree of uncertainty over their estimates cannot be judged as any less than CVM based on the research record to date. Consequently there is no basis for rejecting CVM especially if it is tied with an effort to try to understand how individuals make decisions about infrequent or unfamiliar consumption choices. Early economists, such as Marshall, emphasized the importance of observation of behavior as a key to economic modeling. When that behavior cannot be observed, economists must find ways of understanding how individuals make their choices. The use of CVM, with full recognition of the learning which has accompanied survey research in other social sciences, appears to be the best available basis for understanding individuals' decision making in these areas. This conclusion does not endorse an exclusive reliance on CVM. Moreover, it implies that the surveys should not have an exclusive focus on deriving valuation estimates. Rather, contingent valuation experiments should be regarded as experiments that may permit economists to understand decision processes in areas where unfamiliar or new choices must be made. Theory may help Us understand what ought to be the key elements in these decisions. It can therefore contribute in substantive ways to CVM design. Equally important, more explicit attempts to integrate what is learned from CVM experiments with conventional economic theory should be an essential dimension of future CVM research.

ENDNOTES

Chapter XI

*) Thanks are due Dan Saks and Sharon Smith for some especially helpful discussions of this topic. They are, of course, not responsible for my use (or abuse) of their suggestions. This research was partially supported by the United States Environmental Protection Agency. However, the views expressed are those of the author and not of the Agency.

1) There are important exceptions. For example, Medoff and Abraham (1979) in discussing productivity performance and earnings make a general comment on empirical testing in economics, noting that:

> "Unlike physical scientists, economists typically are not involved in the collection of the data they use, and unlike other social scientists, economists generally avoid having contact with their units of observation. As a result, the proper data for testing numerous important beliefs that many economists hold have not been gathered and the knowledge of those who are likely to really know what is going on has been ignored." (p. 48).

Maital's (1982) recent discussion of the role for psychology in economic modeling brought the Medoff-Abraham's quote to my attention.

2) A simple analytical discussion of the implications of air quality measures for monitoring policies was recently reported by Evans (1984). However, no explicit attention was given to the importance of perceptions in affecting what the author describes as "optimal environmental metrics."

3) See, for example, Linemann (1980), (1981), Krumm (1980), and a large number of others. Bartik and Smith (1984) have recently reviewed the use of hedonic models to evaluate tine role of urban amenities and provide further references.

4) I am grateful to Sharon P. Smith for calling this distinction in the sources of rental information for the calculation of the CPI to my attention.

5) This difference is important because Flinn and Heckman (1983) report, based on the NLS sample, that the categories "unemployed" and "out of the labor force" are behaviorally distinct labor force states. They conclude that:

> "Our empirical results indicate that unemployment and out of the labor force are behaviorally distinct, so that in general it is not legitimate to aggregate the two states into a single unemployment state when analyzing labor market dynamics." (p. 38)

6) Two recent examples include Crandall's (1983) recent critique of air pollution policies where he uses these cost data, along with other cost information, to judge the efficiency of current air quality standards. A second study by Pashigan (1984) uses these data to evaluate the effects of environmental regulation on plant size. Neither directly addresses the prospects for bias with the self-reported data. It should, however, be acknowledged that Crandall assembles several sources of cost data to support his arguments.

7) Mellow and Sider (1983) reported the mean difference in the log of each wage (i.e., log(employer reported) - log(employee reported)) and the variance for this difference. This conclusion is based on testing whether the population mean difference was different from zero. It yielded a t ratio of 7.895.

8) It is not because of the early concerns over the prospects for strategic responses. Strategic behavior does not appear to pose problems with carefully worded questions.

9) A different judgement on the importance of environmental economists' lack of experience with the techniques of survey research that provides an explanation for Carson and Mitchell's (1984) evaluation of the prospects of contingent valuation methods. They suggest that the quality of CVM valuation responses is directly related to questionnaire design, concluding their recent paper on non-sampling errors in contingent valuation research by noting that:

> "... CV (contingent valuation) remains an important and viable method to measure the benefits of many nonmarketed goods. CV is virtually the only method capable of measuring most non-use benefits, such as the value people place on the provision of wilderness areas even when they do not intend to use these areas themselves. While other methods are able to measure use benefits, they are not necessarily superior for that purpose to a well designed and executive CV survey." (p. 21)

10) It is also important to note that there is no reason to believe that the indirect methods' estimates all exhibit the same sampling distributions. The Cummings et al. comparisons of CVM and indirect results treats the travel cost model and hedonic models as equivalent in their accuracy. Each requires quite different assumptions and can be expected to exhibit rather different performance patterns.

11) Strictly speaking, their formulation of the process of developing confidence intervals is confused. Comparison of point estimates of an unknown parameter (an individual's valuation of some environmental amenity) without some information on the nature of

the variation in these estimates and their sampling distributions cannot conclude anything in a formal sense. The authors recognize this and have tried to provide what might be called a judgemental comparison. Such evaluations are inevitably controversial because they require reliance on the analyst's judgement as an alternative to an explicit model of the process leading to each method's estimate, and with it a formal derivation of the properties of each estimator.

12) Maital (1982) made a similar general point in calling for closer coordination between economics and psychology. He noted that the conventional definition of economics leaves out the "why" of the questions (in Knight's terms) that are answered by an economic system (see especially his pp. 15-170.

APPENDIX - Chapter XI
SOME QUIBBLES ON THE CUMMINGS, BROOKSHIRS, SCHULZE SUMMARY OF CVM RESEARCH

There are several points in the Cummings et al. summary of past research that should be clarified.

1. Starting Point Bias

The record on starting point bias seems more clearcut than the Cummings et al. summary appears to suggest. There does appear to be stronger evidence that starting point does matter to CVM estimates using the iterative bidding approach. Tests of the differences in mean option price bids between $25 and $125 starting points in Desvousges et al. (1983) indicated significant differences for all water quality levels. This seems to be consistent with Rowe et al. (1980), and with Mitchell and Carson's (forthcoming) interpretation of the Greenley, Walsh and Young (1981) work. A possible explanation for earlier results where no differences were found between starting points follows from the fairly narrow range in the starting points used for these experiments.

One of the issues that remains unresolved is the relationship between all questioning formats. Here the evidence seems less clearcut than the Cummings et al. report would seem to indicate. For example, the performance attributed to the payment card approach based on recent experiments involves changes in the conditions of what was being elicited (e.g., additional bids were requested after respondents were informed their initial bids would not assure the outcome that had been described to them).

2. Iterative Bidding

The iterative bidding process cannot be paralleled to the learning process that accompanies repeated involvement in an auction process (as is frequently observed in laboratory experiments). Learning time varies, as the authors acknowledge, with the complexity of experimental market process. However, in all cases, market periods involve several minutes each (the time varying with the number of participants) and intervals between these periods, usually for calculations and learning. In some cases, the process can involve over an hour for each experimental trial. By contrast, an hour is often the upper maximum for survey interviews involving a large number of questions. Iterative bidding questions would involve a small fraction Of this time and no mechanism for the individual to learn based on responses to earlier questions. Thus, the parallel to experimental findings with auction mechanisms may be tenuous.

3. The Desvousges, Smith, McGivney Comparative Analysis

Several aspects of the report's summary of Desvousges et al. comparative analysis are inaccurate.

(1) The survey elicited option price, not option value. The interview involved explaining to each respondent the components of total valuation, requesting an option price bid and then asking how much of that response was attributable to anticipated use of the river under improved water quality conditions.

(2) The travel cost model developed as part of the research did consider the opportunity cost of travel time; it did not assume a constant wage rate for all individuals; and it did evaluate the role of model specification, the treatment of on-site time, and the character of the survey data for the travel cost models.

(3) Our comparative analysis was clear on the interpretation of the relationship between contingent valuation and indirect measures of the valuation of water quality. We found that CVM estimates appeared to overstate the travel cost estimates of the value of water quality improvements. This finding was based on our statistical analysis of sixty-nine users' bids and the projected consumer surplus increments for each individual (a total of 94 observations including 16 cases where individuals went to multiple sites). Simple comparisons of the means had the travel cost estimates of consumer surplus falling within the range for the estimated user values across questioning formats. The same was not true with a deterioration in water quality. In this case (where water quality was assumed to deteriorate to a level preventing any use of the river), CVM estimates were substantially less than the travel cost estimates and significantly different (as measured using a hypothesis test of unity for a slope parameter from a regression of the CVM estimate of user value on the travel cost estimate). It was argued that because the travel cost model had to ignore the role of substitute sites, it would be likely that this model would overstate the loss in consumer surplus associated with a water quality reduction hypothesized to lead to the loss of the use of the river's sites for any recreational activities (see Desvousges et al. (1983) pp. 8-16 to 8-18). Thus, the ambiguity in the findings suggested in the Cummings et al. summary of the results is misleading (see their discussion, Chapter 6, p. 163).